

Serial and Single Crystal Studies of
Ligand Binding to Radiation Sensitive
Haem Protein DHP-B

Emily Urban

A thesis submitted for the degree of

Master of Science by Dissertation

School of Biological Sciences

University of Essex

Abstract

Dehaloperoxidase (DHP) is a multifunctional globin and is the first known globin to have a biologically relevant peroxidase function. What makes DHP different from other globins is not just its peroxidase function, but its flexible distal histidine that is located at the opening of the distal pocket, this is essential to its peroxidase function. The main aims of this study are to better understand how and also where dehaloperoxidase binds its ligands, demonstrating different binding modes and pockets to better understand how it switches between its functions. Using both data collected during this project, and processing data from previous experiments, and comparisons of structures obtained from literature, binding modes of different ligands with the isoform dehaloperoxidase-B were explored. Another important question explored in this thesis, is the use of both room-temperature (serial) and cryogenic (single) data collection, to compare whether these two methods produce significantly different results.

Single crystal data were obtained from Diamond Light Source (DLS) and Swiss Light Source, or from the protein data bank. Crystals were grown using the hanging-drop vapour diffusion method, soaked in cryoprotectant and measured at synchrotron sources. Serial data was collected using microcrystals at SACLA, Japan and from DLS. The structures were solved using both *CCP4i2* suite and *Phenix*. The structures displayed different binding modes and pockets present in DHP-B, making bonds with multiple different amino acids present in the binding pocket. The distal histidine (His55) always plays a role in binding due its flexibility. The presence of a hemichrome species is also observed in almost all the structures in this study, including in room temperature, damage-free X-ray Free-Electron Laser (XFEL) structures, ruling out the possibility of the hemichrome species being from the use of cryoprotectant. More research needs to be done to determine whether the hemichrome species has any biological relevance. Comparisons of serial and single crystal studies do show some differences, indicating that temperature measured does have an effect on the structure.

Acknowledgments

I would like to thank my supervisor Dr. Mike Hough for the guidance during this masters degree, and his continued support with the challenges faced during the project. I would also like to thank Hannah Adams for her help with all of the lab protocols and trouble shooting (and there was lots of that!), as well as everyone else in Lab 5.08. Thanks also to all of the team at Diamond Light Source, and the teams at SACLA and Swiss Light Source. `

I would also like to thank my parents for their support, especially financial support which made this masters degree possible for me. Also to Gina Davis, for her emotional support and many coffee breaks! Thanks to Ali Ebrahim for the many beers, they made the awful SACLA dinners bearable, as well as helping me throughout this whole process, this would be two pages without your guidance. To Jaid, for allowing me to complain, rant and rave, and for your continued interest in what I do even if you don't understand a word of it.

Contents

ABSTRACT	2
ACKNOWLEDGMENTS	2
CONTENTS	4
ABBREVIATIONS	6
CHAPTER 1 – INTRODUCTION	7
1.1 INTRODUCTION TO HAEM PROTEINS	7
1.1.1 <i>The Haem Cofactor</i>	7
1.1.2 <i>Functions of Haem Proteins</i>	8
1.2. DEHALOPEROXIDASE	11
1.2.1. <i>Physiological Roles of DHP</i>	12
1.2.2. <i>Structure of DHP</i>	13
1.2.3. <i>DHP Isoforms</i>	16
1.2.4. <i>DHP-B</i>	17
1.2.5. <i>Catalytic Cycle of DHP B</i>	18
1.2.6 <i>DHP ligands</i>	19
1.3. RADIATION DAMAGE	20
1.3.1. <i>Classification of Radiation Damage</i>	20
1.3.2. <i>Reducing Radiation Damage</i>	21
1.3.3. <i>Radiation damage and Temperature</i>	22
CHAPTER 2 - CRYSTALLOGRAPHY METHODOLOGIES	24
2.1. PRODUCING SINGLE CRYSTALS	24
2.1.1. <i>Single Crystal Spectroscopy</i>	24
2.2 SERIAL CRYSTALLOGRAPHY	25
2.2.1 <i>Microcrystallisation</i>	25
2.2.2 <i>Serial Crystallography</i>	25
2.2.3 <i>‘Diffraction before Destruction’</i>	29
2.2.4 <i>Room-Temperature Serial Crystallography</i>	30
CHAPTER 3 - METHODS AND MATERIALS	31
3.1 BACTERIAL TRANSFORMATION AND DNA EXTRACTION	31
3.2 RECOMBINANT PROTEIN EXPRESSION	31
3.3 PROTEIN PURIFICATION	32
3.4 UV-VIS SPECTROSCOPIC ANALYSIS	32
3.5 SDS-PAGE GEL ELECTROPHORESIS	33
3.6 CRYSTALLIZATION	33
3.7 CRYSTAL SOAKING	34
3.8 SINGLE-CRYSTAL MEASUREMENTS	34
3.8.1 <i>Conventional Synchrotron Data collection</i>	34
3.9 SERIAL CRYSTALLOGRAPHY METHODS	34
3.9.1. <i>Batch crystallization</i>	34
3.9.2. <i>Fixed target crystallography</i>	35
3.9.3. <i>Chip Screening</i>	36
3.9.4. <i>SFX Data Processing</i>	36
3.9.4.1 Indexing	37
3.9.4.2 Merging	38
3.9.4.3 Data Quality Statistics	39
3.9.4.4 Producing an mtz file for molecular replacement and refinement	39

3.9.5. SSX Data Processing	39
3.10 X-RAY DATA	40
3.11 MOLECULAR REPLACEMENT AND REFINEMENT	40
CHAPTER 4 – SINGLE CRYSTAL DHP CRYSTALLOGRAPHY	42
4.1 INTRODUCTION	42
4.2 PROTEIN PURIFICATION	42
4.3. CRYSTALLISATION	43
4.4. CRYSTAL SOAKING	43
4.4.1 <i>Ligand Bound Structure of DHP-B with 7-bromoindole</i>	44
4.5 DHP-B CRYSTAL SOAKING EXPERIMENTS AT SLS	46
4.6 COMPARISONS FROM LITERATURE	47
4.7. DISCUSSION	49
CHAPTER 5 - SERIAL DHP CRYSTALLOGRAPHY	51
5.1 BATCH CRYSTALLIZATION	51
5.2 SYNCHROTRON SERIAL CRYSTALLOGRAPHY (SSX) AT DIAMOND LIGHT SOURCE	52
5.2.1 <i>Processing of SSX data</i>	53
5.3 SERIAL FEMTOSECOND CRYSTALLOGRAPHY (SFX) AT SACLA	55
5.3.1 <i>Measured SFX Data for DHP -7-bromoindole Complex</i>	56
5.3.2 <i>Processing SFX Data</i>	57
5.4 SFX STRUCTURE OF DHP WITH PERACETIC ACID (FORMATION OF COMPLEX L)	68
5.4.1 <i>Comparison of Ferric DHP-B and DHP-B and Peracetic Acid</i>	71
5.5 SFX STRUCTURE OF OXYFERROUS DHP	73
5.5.1 <i>Comparison of DHP Oxyferrous – Room Temperature (SFX) and Cryogenic</i>	80
5.6. DISCUSSION	82
5.6.1. <i>DHP Serial Crystallography</i>	82
5.6.2 <i>Comparison between Ferric DHP and Peracetic Acid soaked DHP</i>	83
5.6.3 <i>Oxyferrous DHP-B</i>	84
5.6.4. <i>Comparison between SFX room temperature and cryostructure oxyferrous DHP-B</i>	85
CHAPTER 6 - GENERAL DISCUSSION	88
6.1 INTRODUCTION	88
6.2 SERIAL CRYSTALLOGRAPHY	89
6.3 LIMITATIONS OF THIS STUDY	89
BIBLIOGRAPHY	91

Abbreviations

DHP – Dehaloperoxidase

Kpi – potassium phosphate buffer

SX – Serial crystallography

SFX – Serial Femtosecond crystallography

SSX – Synchrotron Serial crystallography

PDB – Protein Data Bank

SDS-PAGE – Sodium Dodecyl Sulphate Polyacrylamide Gel
Electrophoresis

UV-vis – Ultraviolet-visible

SLS – Swiss Light Source

DLS – Diamond Light Source

XFEL – X-ray Free-Electron Laser

Amp – ampicillin

IPTG – isopropyl β -D-1-thiogalactopyranoside

LB – Luria-Bertani Broth

PEG – Polyethelen Glycol

TRIS – Thrisaminomethane

rpm – revolutions per minute

kDa – kilodalton

V – Voltage

RMSD – Root Mean Square Deviation

MES - 2-(*N*-morpholino)ethanesulfonic acid

DCP - 2,4-dichlorophenol

Chapter 1 – Introduction

1.1 Introduction to Haem Proteins

Haem proteins are among the most studied families of proteins (Reedy, Elvekrog and Gibney, 2008). They are a vast family of metalloproteins with a large range of functions. The haem group present in these proteins gives rise to their function and allows the protein to carry out reactions that are vital for cellular processes and other ligand binding properties. The most well-known haem proteins are myoglobin and haemoglobin, both known for their oxygen storage and transportation, and were also the first proteins to be crystallised. Haem proteins have a characteristic red or brown colour, depending on the oxidation state of the cofactor, due to their haem chromophore.

The functional versatility of haem groups is due to the protein's ability to tune haem reactivity, based on the number of protein-donated axial ligands to iron, the nature of these ligands, where the haem is located in the protein which therefore dictates its accessibility to external ligands, as well as a number of other properties of the haem-binding site (Paoli, Marles-Wright and Smith, 2002).

1.1.1 The Haem Cofactor

The haem cofactor is a prosthetic group and an aromatic polycyclic molecule. The haem cofactor is also known as protoporphyrin IX. This is made up of a central iron, either ferric or ferrous in resting state, and ferryl in reactions, non-covalently bound to the protein, surrounded by four pyrrole molecules (forming the backbone). These are joined together by methylene bridges forming a so-called tetrapyrrole ring (Schneider *et al.*, 2007) Haem is considered an iron-bound porphyrin, as when the tetrapyrrole binds to a metal it is considered a porphyrin.

There are many types of haem prosthetic groups, with haem b and c being the most common. There is also a, d, f and o type haems (shown in Figure 1.1). Each different type of haem has different substituents in the outer ring, giving them their unique

functional roles including key cellular process and their ability to act as an enzyme in catalysis (Reedy, Elvekrog and Gibney, 2008)

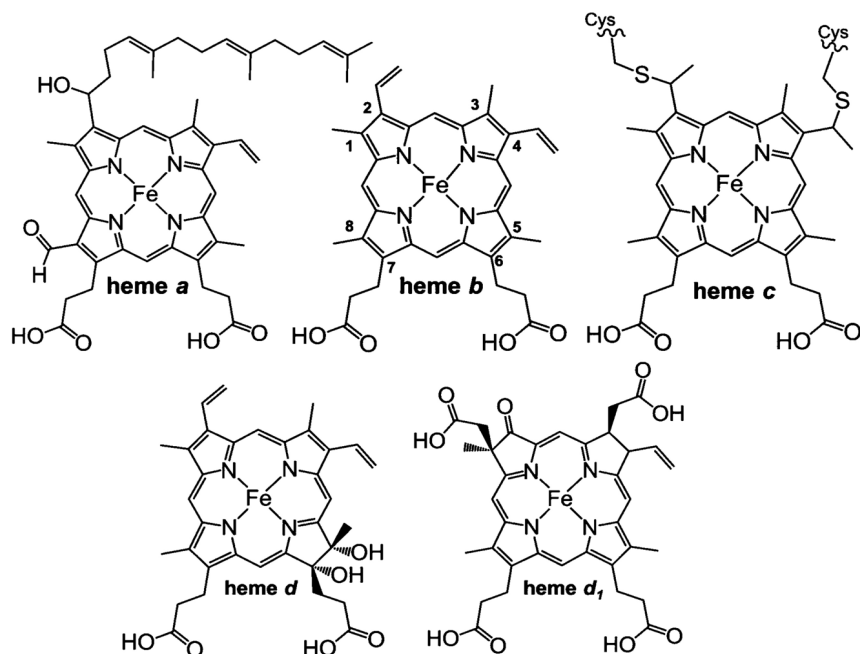


Figure 1.1: structures of the different types of haem groups, labelled a to d (Liu et al 2014). Each different haem group has unique functional roles, including in catalysis and key cellular processes.

1.1.2 Functions of Haem Proteins

The different haem prosthetic groups allow haem proteins to have a vast array of functions. When the iron is in a ferrous redox state, it can reversibly bind gaseous ligands enabling the protein to perform roles of oxygen storage and transport, such as in haemoglobin and myoglobin. Cytochromes a, b, and c are haem proteins that perform an electron transfer function and are found in electron transport chains. Haem proteins also have roles as gas sensors for gases such as oxygen and nitric oxide and cell redox state, as well as regulating circadian rhythms (Girvan and Munro, 2013) They can also function as enzymes for catalysis, such as cytochrome p450s, cytochrome c oxidase and peroxidases (Ortiz de Montellano, 1987)

A widespread role of iron in haem proteins in biological systems is producing iron

intermediates, known as the ferryl haem species. Haem enzymes that use this include all cytochrome p450s and haem peroxidases. Despite peroxidases versatility, they all perform catalysis in the peroxidase cycle in a similar way. There are two ferryl intermediates produced, known as compound I and compound II (Moody and Raven, 2018). The most well-studied peroxidase, horseradish peroxidase, has a different compound I species, containing a ferryl haem species as well as a porphyrin π -cation radical. Other peroxidases compound I contain just a porphyrin π -cation radical (Morita *et al.*, 1988)

1.1.3 Peroxidases

In the absence of exogenous electron donors, peroxidases can slowly degrade hydrogen peroxide by a pseudo-catalase cycle which involves several redox intermediates known as compounds I, II and III (Vlasits *et al.*, 2010). In the first step, peroxide removes one electron from the iron, and a second from the porphyrin to generate porphyrin π cation radical (Dophin *et al.* 1971). This has a green colour, with distinct spectral characteristics very different to the brown-red colour of the resting haem enzyme. In the second step, a substrate molecule moves one electron to compound I, reducing the porphyrin cation radical giving the red colour of compound II. In the last step, a second substrate reduces Fe^{4+} back to Fe^{3+} . The role of the peroxidase determines the nature of the substrate. The structures of the different intermediates and spectra for the different intermediates are shown in Figure 1.2 and 1.3.

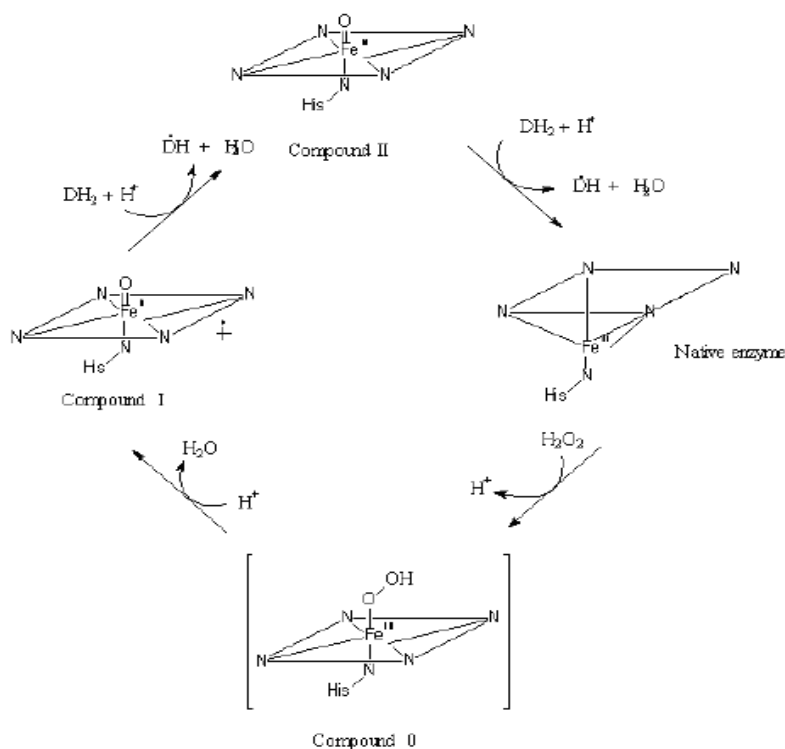


Figure 1.2: The catalytic cycle of Horseradish Peroxidase, showing the structures of Compound I, Compound II and native state (Longu et al 2004)

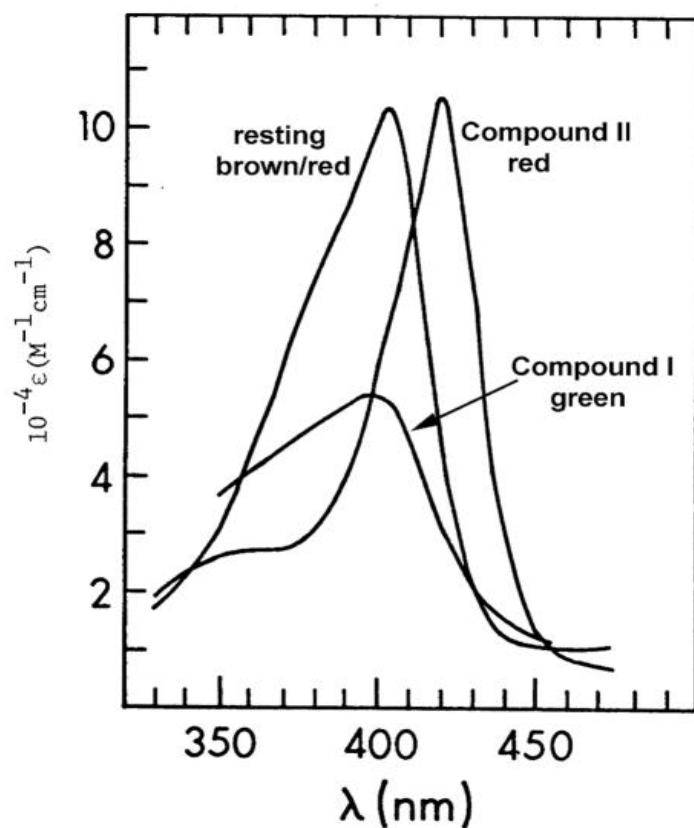


Figure 1.3: UV-Vis Spectra of the three intermediates in HRP Catalysis (Dunford 2010).

Peroxidases are divided into two main families, haem and non-haem peroxidases depending on whether they contain an iron cofactor. Haem peroxidases are split into four superfamilies, Peroxidase-Catalase, Peroxidase Cyclooxygenase, Peroxidase-chlorite dismutase, and Peroxidase-peroxygenase. These are shown in Figure 1.4 (Neyadi, Rauf and Ashraf, 2018). These enzymes are single polypeptide chains of $\approx 30,000\text{--}40,000$ Da, with a single haem group attached to the protein via iron ligation to a His residue (Poulos, 2014). All of the peroxidases crystal structures show that the overall core helical and active site structures are well conserved (Sivaraja *et al.*, 1989).

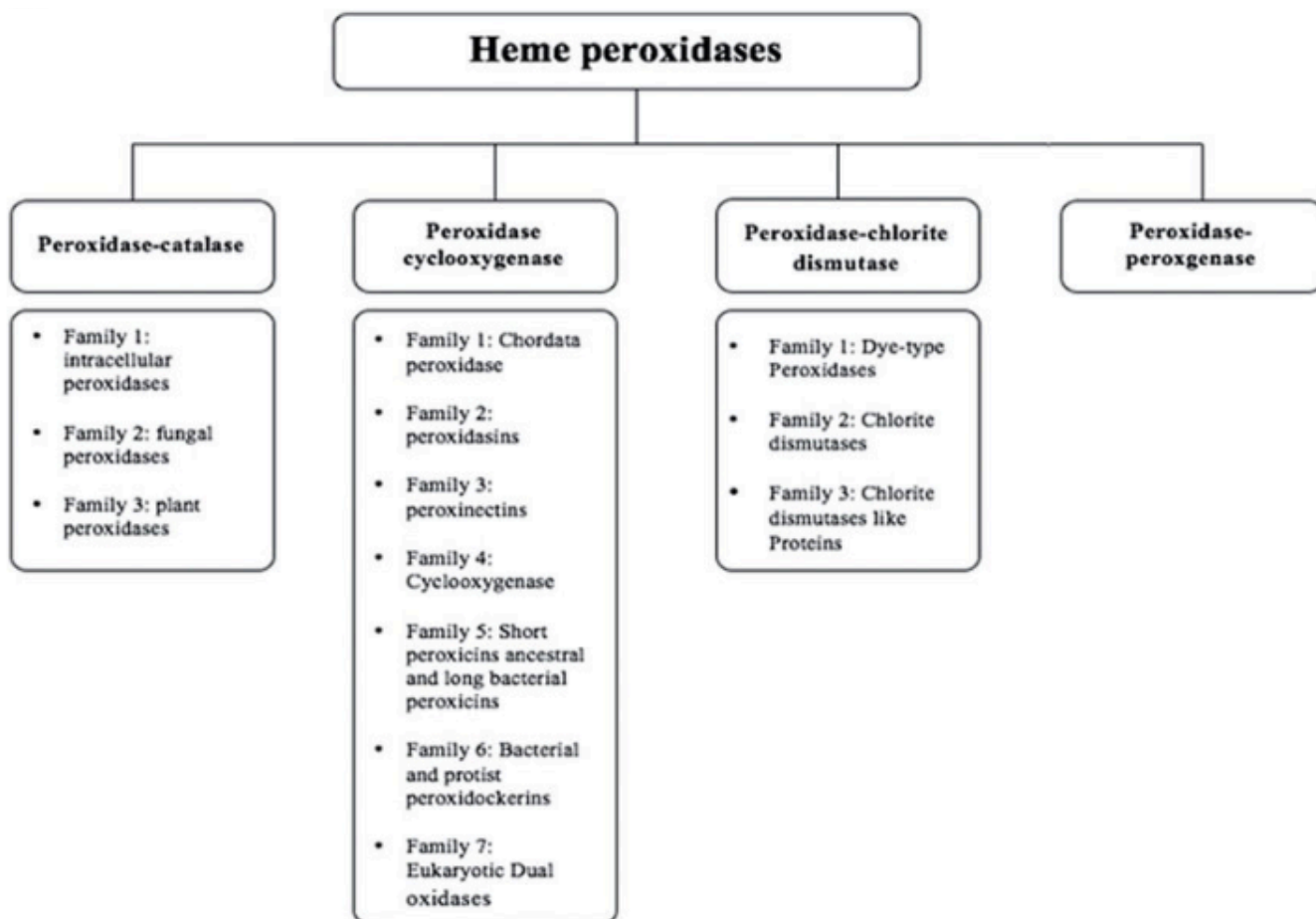


Figure 1.4: The four haem-peroxidase superfamily classification and their subfamilies. (Neyadi et al 2018). Each family arose differently through evolution, and have different overall folds, active site architecture and enzymatic activities.

The haloperoxidase family of peroxidases are abundant mainly in fungi. There are two different groups of haloperoxidase family, ones with prosthetic haem groups and nonhaem haloperoxidases. These two groups are phylogenetically unrelated, and form two different gene families. Haem-thiolate haloperoxidases have a protoporphyrin IX prosthetic group, and catalyse the oxidative degradation of chloroperoxidases (Osborne *et al.*, 2006)

1.2. Dehaloperoxidase

Dehaloperoxidase (DHP) is a haem-containing peroxidase, globin and the coelomic haemoglobin of *Amphitrite ornata*, a marine annelid found in Cobscook Bay, the Gulf

of Maine and in the north west Atlantic Ocean (Lebioda *et al.*, 1999). DHP is the first globin to have an identified, biologically relevant peroxidase function (de Serrano *et al.*, 2007). DHP is multifunctional, functioning as a peroxidase, haemoglobin, peroxygenase, oxidase and oxygenase. This is highly unusual and unique, as haem-proteins tend to only have one key function, with many having only 'moonlighting' alternative functions. Peroxidase activity in DHP can be initiated from both ferric and oxyferrous oxidation states, binding molecular oxygen in ferrous state.

1.2.1. Physiological Roles of DHP

Amphitrite ornata is a terebellid polychaete found in shallow mud flats and marine estuaries. These ecosystems have many other polychaete worms such as *Notomastus lobatus*, which contaminates the sediments with mono-, di- and tribromophenols and hemichordata like *Saccoglossus kowalewski*, that secrete halogenated aromatic compounds that are highly toxic. These organisms all secrete these as a repellent for predators as well as preventing attacks by fungi (Lincoln *et al.*, 2005). A number of these species also produce brominated phenols, pyrroles, and indoles. DHP has evolved to degrade and detoxify these compounds, which allows *A.ornata* to live in these ecosystems (Chen *et al.*, 1991). DHP levels are very high, representing 3% of the soluble protein in *A.ornata*. DHP catalyses the oxidative dehalogenation of polyhalogenated phenols in the presence of hydrogen peroxidase at a rate 10 times faster than all known halohydrolases of bacterial origin, this reaction is shown in Figure 1.5 (LaCount *et al.*, 2000). All of DHP's functions could potentially protect *A.ornata* from the other organisms in its ecosystem.



Figure 1.5: Trihalophenols oxidized by hydrogen peroxide producing dihaloquinones, catalyzed by DHP

Genetic analysis shows that DHP is encoded by a gene that has a high sequence homology to other globins in the annelid family (Bailly *et al.*, 2007) and shares structural characteristics with other peroxidases as well as enzymatic similarities. DHP is thought to have evolved from an ancient oxygen transport and storage globin, with haemoglobin phylogeny revealing a common ancestry across species extending back 1.8 billion years (Hardison, 1998) therefore retaining its oxygen carrying function while also being able to act as a peroxidase (Du *et al.*, 2011). DHP has two different isoforms, DHP-A and DHP-B with 96% amino acid identity, with DHP-B being significantly more reactive.

1.2.2. Structure of DHP

The X-ray crystallographic structure of DHP was first determined at 1.8 Å by LaCount *et al* (2000). This revealed that DHP contains the globin fold, but its haem moiety is located 1.5 Å deeper in the protein than in myoglobin. DHP is a homodimer, with two identical subunits of ~15.5 kDa in the asymmetric unit, each containing a haem protoporphyrin XI cofactor and eight alpha-helices (de Serrano *et al* 2007). The overall structure of the DHP dimer is shown in Figure 1.6. The subunits are connected by salt bridges and one hydrogen bond between Asp72 and side chain groups of Arg122 and Asn126 (Lebioda *et al.*, 1999) shown in Figure 1.6. There are

hydrophobic interactions between Val74 residues of the two subunits. There is one cysteine in DHP, located close to the interface. The distal cavity in DHP is as hydrophobic as in other globins, and does not have the polarising auxiliary arginine present in other peroxidases. Peroxidases need this arginine, as it stabilises the charge in the distal cavity, that becomes negative when the distal histidine is functioning as an acid/base in proton transfer (LaCount *et al.*, 2000).

DHP was originally identified as a dimer (Weber *et al.*, 1977). The crystal structures of DHP show a dimer in the unit cell, but DHP has a significantly smaller dimer interface (450 \AA^2) than the typical values of $1200\text{-}2000 \text{ \AA}^2$ for other dimers. Thompson *et al* discovered using fast protein liquid chromatography and small-angle X-ray scattering, that DHP in fact is found primarily monomeric in solution, with some detectable levels of dimer, around 10% (Thompson *et al.*, 2011).

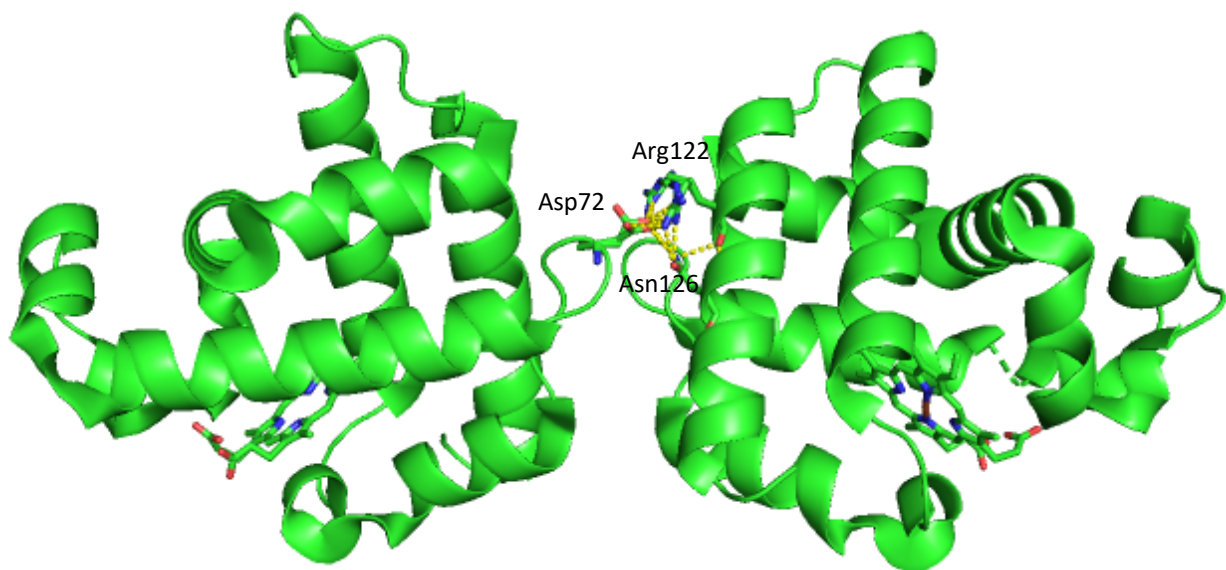


Figure 1.6 Overall structure of DHP-B dimer at 1.58 \AA , pdb code 3IXF (de Serrano *et al* 2010). Amino acids involved in salt bridges are labelled, with the bonds shown by yellow dashed lines.

The first crystal structure of substrate-bound DHP showed that 4-iodophenol binds in the distal pocket of the DHP isoform DHP-A, with no coordination to the haem iron. This internal binding site distinguishes DHP from both other globins and haem

peroxidases (de Serrano *et al.*, 2007). Para-halogenated phenols can occupy this internal binding site. However, most active substrates of DHP 2,4,-di and 2,4,6-trichlorophenols do not readily enter this distal pocket (de Serrano *et al.*, 2007). This suggests this internal binding site is not the active site. There is the potential that both external and internal small-molecular-binding sites exist in DHP. This may explain how DHP switches between peroxidase and haemoglobin activities, which is currently unknown (Chen *et al.*, 2009).

The main difference between DHP and other globins like myoglobin is the distal histidine, which is further from the haem iron in DHP. Instead, the distal histidine is located at the entrance of the distal cavity, making it more suitable for the heterolytic cleavage of hydrogen peroxide and facilitates the movement of the histidine out of the distal pocket to make room for organic substrates (Lebioda 2000), which is more characteristic of peroxidases. While peroxidases tend to have at least one polar amino acid participate in activation of bound hydrogen peroxide, the distal histidine in DHP is the only polar amino acid in the distal pocket (Plummer, Matthew and Franzen, 2013). The distal histidine His55 has been observed to take on 'open' and 'closed' conformations. When in 'closed' conformation, His55 points inside the distal pocket and interacts with the sixth ligand (water, oxygen) of haem iron (Chen *et al.* 2009). The 'open' conformation is when His55 is rotated to a solvent-exposed position, observed when corresponding 5-coordinate haem iron either in deoxyferrous or when inhibitors are present in distal pocket (Francesco *et al.*, 2010)

The haem proximal side in DHP is also different from other globins. The proximal histidine (His89) is shifted in the sequence by two residues, and consequently the main chain loop position differs and the plane of imidazole moiety is rotated when compared to myoglobin (LaCount *et al.* 2000). The structural differences between the positions of the proximal histidine of DHP and myoglobin are represented in

Figure 1.7.

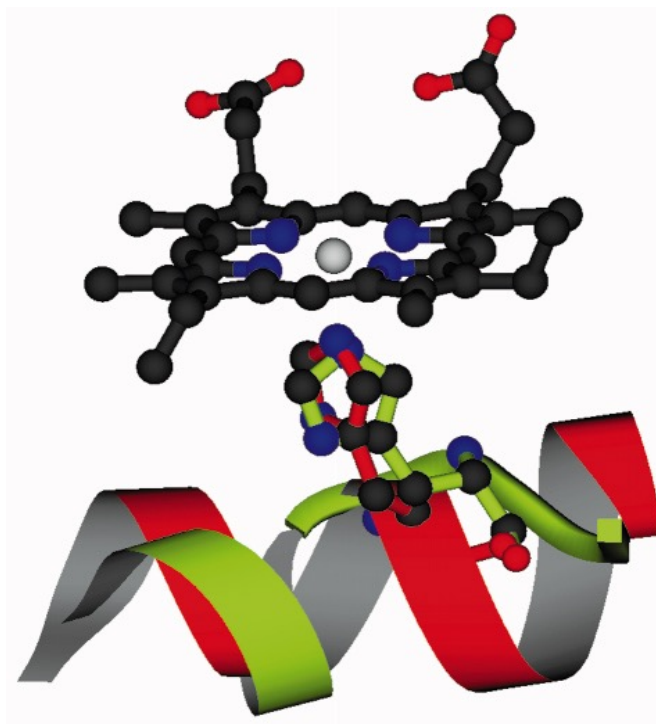


Figure 1.7: Molscript representation of the least-squares superposition optimizing haem overlap, showing the positions of the proximal histidine in DHP (green) and myoglobin (red). The main chain with histidine is in a coil conformation in DHP, while myoglobin is a helical conformation (Lebioda 2000)

1.2.3. DHP Isoforms

There are two isoforms of DHP, DHP-A and DHP-B, both with 137 amino acid residues which have 96% amino acid identity, differing by only 5 amino acids around its active site, despite these differences they are both able to catalyse the oxidation of 2,4,6-trihalogenated phenol in the presence of co-substrate H_2O_2 . The two isoforms of dehaloperoxidase are encoded by two sperate genes. Although they have a high amino acid identity, there are significant spectroscopic and mechanistic differences between them (D'Antonio, J., D'Antonio, E., Thompson, M.K., Bowden, E.F., Franzen, S., Smirnova, T. and Ghiladi, 2010), this includes the fact that DHP-B has four times higher peroxidase activity than DHP-A for 2,4,6-tribromophenol, having highest peroxidase activity of any known naturally occurring haemoglobin

(Franzen *et al.*, 2016). It is also known that DHB forms a compound RH intermediate which exhibits different reactivity than DHP-A. DHP-B also has a greater extent of substrate inhibition (Feducia *et al.*, 2009). The percentage of each DHP isoform is unknown, but current evidence suggests that DHP-A is the more dominant form of DHP in *A.ornata* (LaCount et al 2000).

1.2.4. DHP-B

Currently, most of the focus on DHP has been on the DHP-A isoform. There has been less research into the mechanism or spectroscopic properties of DHP-B compared to its other isoform. DHP-B differs from DHP A at five positions: I9L, R32K, Y34N, N81S and S91G. The crystal structure of DHP-B is shown in Figure 1.8.

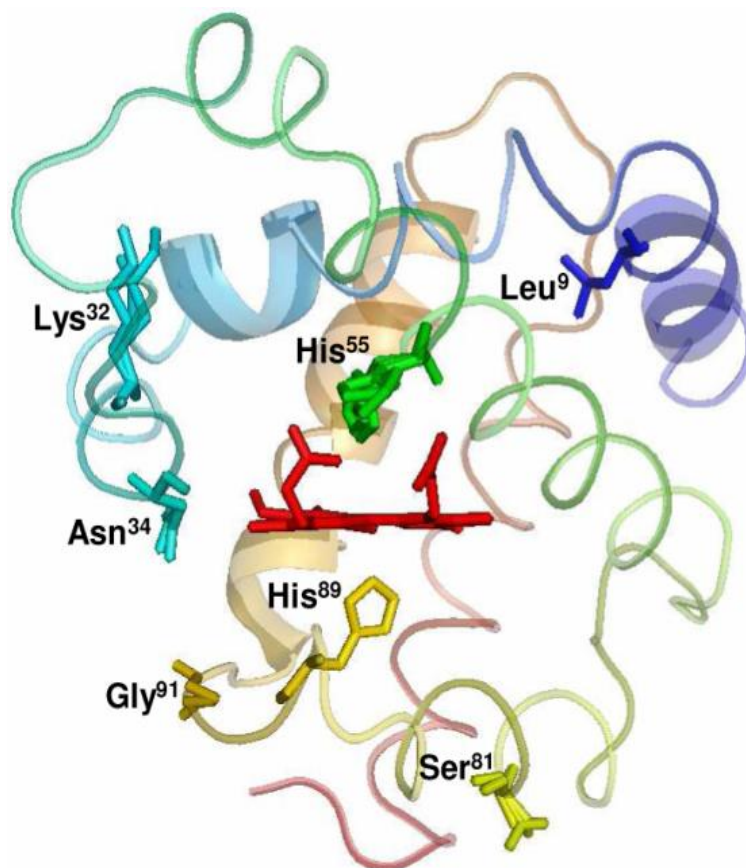


Figure 1.8: Crystal structure of DHP-B (PDB: 3ixf). The location of the residues that are different from DHP-A are labelled as well as the proximal (His89) and distal (His55) histidine's. (D'Antonio et al 2010)

As mentioned previously, DHP-B is twofold to fourfold more active than DHP-A, depending on the substrate. DHP-B also demonstrates a greater substrate inhibition than DHP-A. however, both isoenzymes have the same distal histidine conformational flexibility which is key to their peroxidase activity (de Serrano *et al.*, 2010)

1.2.5. Catalytic Cycle of DHP B

Dehaloperoxidase has four different kinds of reaction, peroxidase, peroxygenase, oxidase and oxygenase (Malewschik *et al.*, 2019). DHP catalyses the oxidative degradation of 2,4,6-trihalogenated phenols to 2,6-dihalo-1,4-benzoquinones in the presence of hydrogen peroxide, as previously stated. DHP reacts with hydrogen peroxide to yield Compound ES, an iron(IV)-oxo haem centre with an amino acid radical (Feducia et al 2009). This intermediate has a catalytic role in the oxidising of co-substrate 2,4,6-trichlorophenol (TCP). In the absence of co-substrate, compound RH is created, unique to dehaloperoxidase. The overall two-electron co-substrate oxidation of TCP is thought to be done in one one-electron step, suggesting that co-substrate binding of DHP occurs in an external binding site (Davis et al 2009).

Dichloroquinone (DCQ), the product of trihalophenol dehalogenation, reacts with Compound ES leading to oxyferrous DHP. DCQ reacts with ferric DHP, also leading to oxyferrous DHP formation. Oxyferrous DHP catalyses the oxidative dehalogenation of TCP in the presence of H₂O₂. 4-bromophenol, a proposed co-substrate, is known to bind in the distal cavity of the haem active site, (Lebioda *et al.*, 1999) is an inhibitor of 2,4,6-trichlorophenol dehalogenation. This proposed catalytic cycle of DHP is outlined in Figure 1.9.

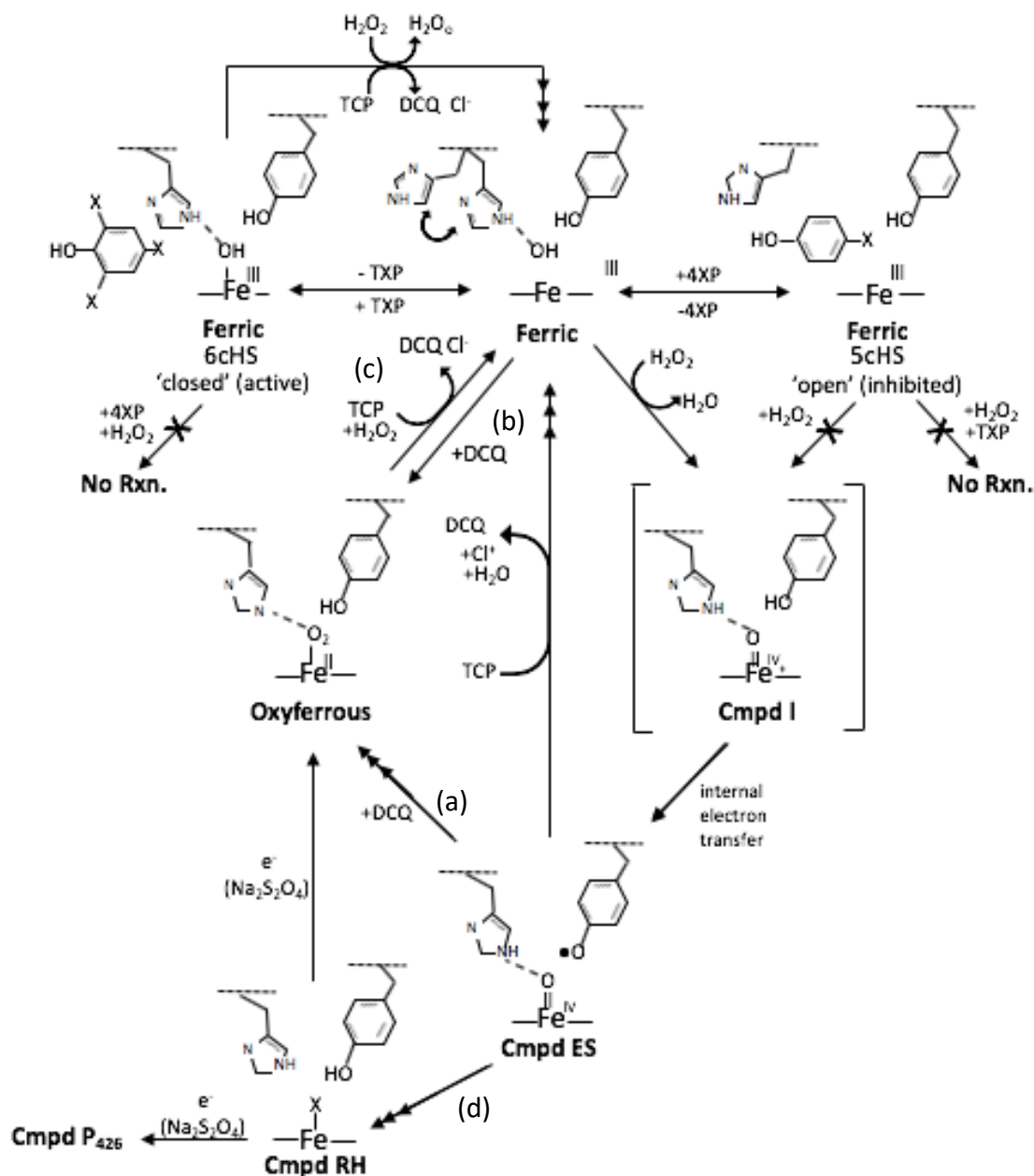


Figure 1.6 Proposed catalytic cycle of DHP

- (a) DCQ (dichloroquinone), the proposed product of trihalophenol dehalogenation, reacts with oxidant compound ES leading to oxyferrous DO
- (b) DCQ reacts with ferric DHP also leading to oxyferrous DHP
- (c) oxyferrous DHP catalyses the oxidative dehalogenation of TCP in the presence off hydrogen peroxide
- (d) In the absence of cosubstrate, there is the formation of compound RH unique to DHP

1.2.6 DHP ligands

DHP has multiple substrate binding sites, and it is proposed that the oxidation state can affect which ligands bind, and is the hypothesis of this study. 5-bromineindole appears to bind in the ferric structure directly above the haem, where the distal histidine is 'swung out' in the open conformation. 5-nitrophenol is thought to also bind in a similar way. Other ligands include 4-bromophenol and 5-bromophenol.

5-bromoindole binds in the distal pocket and forms a hydrogen bond with the distal histidine (Barrios *et al.*, 2014). The distal histidine His55 is in a 'closed' conformation, allowing the 5BR to bind in the pocket. 5BR does not form any bonds with iron as it's bound higher up in the pocket.

1.3. Radiation Damage

The creation of intense undulator beamlines from second and third generation synchrotron sources has caused a previous problem in crystallography to remerge: X-ray radiation damage. This is a problem even at cryogenic temperatures. In room temperature experiments, this may mean having to use multiple crystals to obtain a complete dataset. The cause of this damage is the energy lost by the beam in the crystal, due to either total absorption or inelastic scattering of X-rays as they pass through the crystal (Garman, 2010)

1.3.1. Classification of Radiation Damage

Systemic studies of radiation damage have shown two separate indicators of damage as a function of dose: global and specific damage. Global damage results in a lowering of measured reflection intensities (especially at high resolution), expansion of the unit-cell volume, increasing values of the measure of the internal consistency of the data (R_{meas}), and increase in the scaling of B factors for the data and the atomic B values of the refined structure, rotation of the molecules within the unit cell, and sometimes an increase in mosaicity. Some visible differences can also be seen, such as colour changes arising from changes in redox state, such as bleaching. Following irradiation, the sample can be warmed and bubbles of hydrogen are emitted and the sample may be discoloured (Meents *et al.*, 2010)

Specific damage occurs at far smaller doses than global damage, with the Garman limit of dose being 20-40 MGy for useful data (Owen, Rudino-Pinera and Garman, 2006). Active sites and redox centres are known to be more sensitive to damage by photoreduction by x-ray beams (Weik *et al.* 2000). Specific structural damage to

covalent bonds can be seen to occur in a specific order in many proteins, changes to metal sites occur first, and then disulphide bridges elongate and break, then glutamates and aspartic acids are decarboxylated, tyrosine residues lose their hydroxyl group and subsequently the carbon-sulphur bonds in methionines are cleaved (Ravelli and Garman 2006). These specific damages can be seen in Figure 1.10.

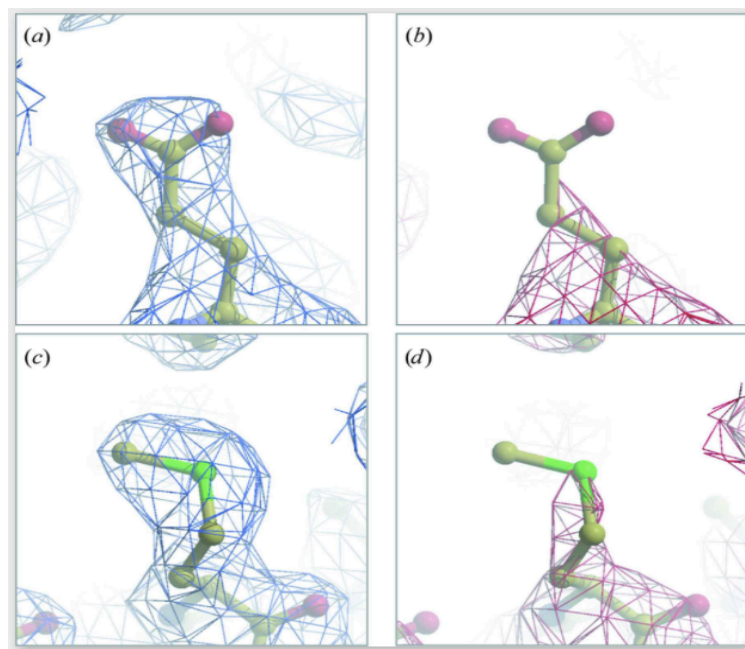


Figure 1.10: Specific radiation damage inflicted on apoferritin cryocooled crystal during sequential data sets. (A) Glu63 after a dose of 2.4MGy. (B) Glu63 after 50MGy (C) Met96 after a dose of 2.5 MGy (D) Met96 after 50MGy (loss of electron density around disordered atoms) (Garman & Owen 2006)

1.3.2. Reducing Radiation Damage

Initially, a simple solution to radiation damage was to collect cryocooled crystals at 100K, as radiation damage accumulates slower at lower temperatures, creating an almost 100-fold extension of the crystals lifetime, reducing secondary radiation damage and usually resulting in higher resolution and better quality data (Garman 2003). Diffraction data is now almost always collected at cryogenic temperatures. However, cryocooling does not eliminate X-ray damage.

Secondary damage is caused by the highly reactive radiolytic species that propagate through the crystal reacting with macromolecules causing further damage, as electrons are mobile at 100K while other radicals are not. Manipulating the physical and chemical environment of the crystal can reduce the amount of secondary damage but is less favorable, such as by collecting data from flash-cooling crystals at liquid helium temperatures (<20K) (Chinte *et al.*, 2007). It is also possible to reduce secondary damage by using a different cryoprotectant agent. Another option would be to incorporate free radical scavengers, such as ascorbate, into the crystal lattice by soaking or co-crystallization to reduce the interaction with the protein (Murray and Garman, 2002)

1.3.3. Radiation damage and Temperature

As stated previously, crystallography is almost always performed at cryogenic temperatures (100K) to reduce radiation damage when collecting data. However there is an increased recognition that collecting X-ray diffraction data at cryogenic temperatures could mask alternate conformation states that are accessible to the protein at room temperature (Fraser *et al* 2009). While cryogenic temperatures have been shown to obtain high-resolution structures, it does not show the sample's structural distributions or the dynamics present at physiological or room temperatures. Therefore, Cryogenic temperatures alter the equilibrium of protein conformations, introducing a small amount of bias due to the cryogenic temperatures minimally changing the overall protein backbone fold (Fraser *et al.*, 2009)

With room temperature being favourable, there is a need to 'outrun' radiation damage to get undamaged data from radiation-sensitive systems. It has recently been discovered that radiation damage can be somewhat outrun at synchrotrons at room temperature, as the damage response is dose-rate dependent (Warkentin *et*

al., 2013) and damage can be completely avoided with 10fs pulses at XFELs to give damage free structures.

Chapter 2 - Crystallography Methodologies

X-ray crystallography is a powerful tool for structural determination of the three-dimensional structure of proteins at atomic resolutions. There are multiple different techniques, such as vapour-diffusion, hanging drop vapour-diffusion, sitting drop, batch and dialysis methods (McPherson, 1999). Producing suitable crystals can be difficult, with optimal conditions being hard to predict. Many different conditions must be tested, making it very time consuming and a rate-limiting step.

2.1. Producing Single Crystals

The first step, as in all crystallography is producing a large enough crystal, with a regular structure with minimal imperfections. This is often the most difficult part, time consuming part.

The principle of crystallisation is to take a sample of a very high concentration and induce it to come out of solution and into the solid phase. If this occurs too quickly, precipitation will happen, but under the right conditions crystals will grow (Luft *et al.*, 1994) 'Crystal screens' are often used, with each one usually containing 50+ solutions, often 96 for robotic screening, varying in precipitant, buffer, pH and salt. This is known as a sparse matrix, and can be set up using vapour diffusion, hanging drop vapour diffusion and sometimes dialysis, at both 4°C and room temperature (Hampel *et al.*, 1968)

To improve crystal size, various techniques can be used. These include seeding, altering protein concentration, or altering the temperature. Crystals can be as small as 10 microns in size at microfocus beamlines, where the X-ray beam can be reduced to 1-5 microns in size (Grimes *et al.*, 2018)

2.1.1. Single Crystal Spectroscopy

Single-crystal x-ray diffraction is the oldest, best established, most precise method

of crystallography. It involves a beam of X-rays hitting a single crystal, producing scattered beams, which are measured by the detector, creating a diffraction pattern. These are recorded at different angles as the crystal is rotated.

Crystals that are big enough and contain the macromolecular subunit of interest are mounted in a capillary tube at room temperature or in a small loop in a stream of cold nitrogen gas at 100K. Data collection at cryogenic temperatures reduces radiation damage, and a complete data set can be collected from a single crystal.

2.2 Serial Crystallography

2.2.1 Microcrystallisation

The issue of growing large crystals for x-ray crystallography experiments can be a huge obstacle, as growing suitably large and diffracting crystals can be extremely difficult. Radiation damage and crystal size are also linked, with damaging radiation depending on dose. When large crystals are not available, the use of multiple small crystals is common, with the data sets being scaled and merge into one set.

Microcrystallography can be carried out at several beamlines at synchrotrons, delivering monochromatic beams of 10 microns in diameter, usual with 10^{12} photons s^{-1} (Smith, Fischetti and Yamamoto, 2012).

A typical microcrystallography experiment is performed by mounting a sample loop with one or more crystals. These are centred in the beam to collect partial data sets. This can be repeated with multiple sample loops. Microcrystals are also needed for serial crystallographic work.

2.2.2 Serial Crystallography

Serial crystallography is when structural information is obtained from small crystals, illuminated by either synchrotron or 'XFEL's. Crystal structures of biological macromolecules are usually limited by the availability of larger-sized crystals, and by the risk of radiation damage during data collection (Gati *et al.*, 2013) Advanced

third-generation synchrotrons are able to produce micrometre-sized (less than 10 micrometres) high-flux X-ray beams which can focus on small crystals, and so can determine crystal structures from crystals with volumes less than $1000\mu\text{m}^3$ (Cusack *et al.*, 1998)

When using microcrystals, its often not possible to locate the crystals using optical microscopy, due to them being too small, optical distortions or the crystal being embedded in an opaque matrix. Single-crystal diffraction patterns are kept, indexed and merged to a form a three-dimensional set of reflection intensities.

Serial crystallography has different possible sample delivery methods. The usual method for serial femtosecond crystallography (SFX) uses a gas-focused liquid jet of the crystal suspension, flowing across the x-ray beam. The jet diameter is low, ranging from 500nm to a few micrometres, giving a low background. The detector frames after each X-ray pulse. Figure 2.2 illustrates this equipment at Linac Coherent Light Source, where they use the LCLS Coherent X-ray Imaging instrument (CXI). It is equipped with Cornell-SLAC Pixel Array Detectors (CSPADS). The CSPAD supports the 120-Hz readout rate needed to measure x-ray pulses from LCLS (Philipp et al 2011). This set up is shown in Figure 2.1

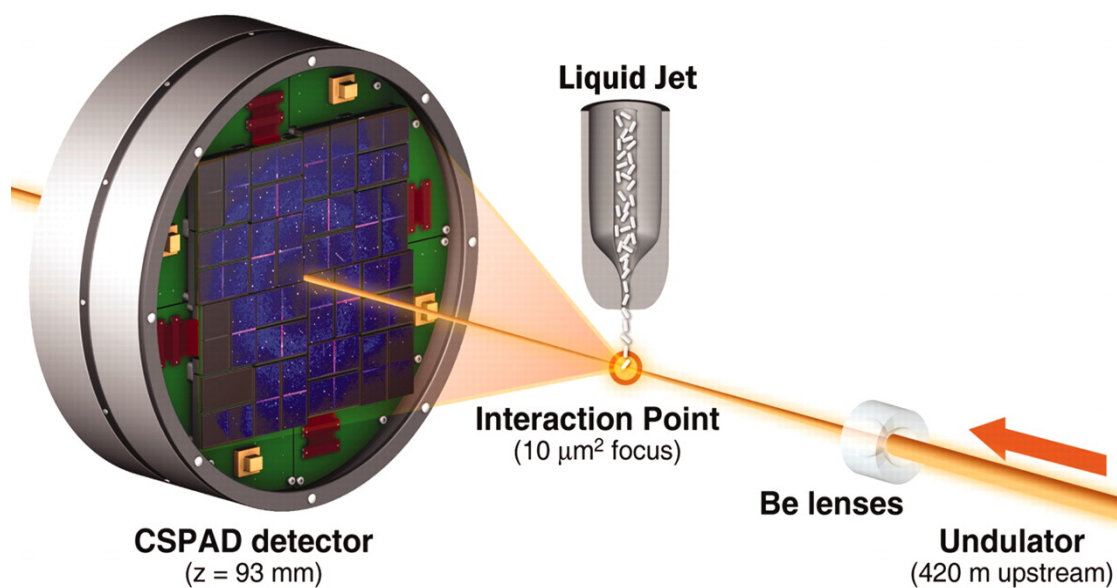


Figure 2.1: Serial Femtosecond crystallography (SFX) instrument. The diffraction patterns are collected on the CSPAD detector. (Boutet et al 2012)

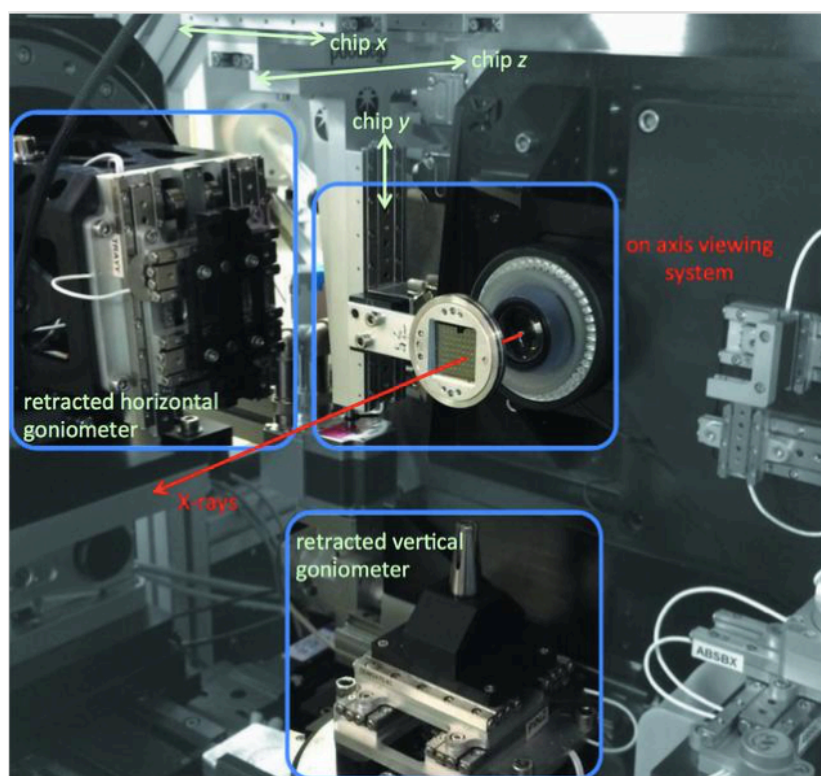


Figure 2.2: Instrument for fixed target serial crystallography at beamline I24 at Diamond Light Source. (Owen et al 2017) The chip is mounted to a high-speed, high-precision xyz stage mounted at the same position, allowing data collection from all 25,600 positions on a chip in less than 10 minutes.

Fixed target crystallography offers high hit-rates with a small amount of sample consumption. This approach can be used at both synchrotrons and XFELs. The samples are pipetted onto silicon nitride 'chips', capable of holding >25,000 crystals. These chips are mounted on a high-speed, high-precision xyz stage mounted at the sample position (Owen et al 2017). This set up is used at Diamond Light Source at beamline i24 and is shown in Figure 2.2

The 'Chip' is a silicon nanofabricate crystallography chip, which can hold the thousands of protein crystals under ambient conditions at defined positions. These chips can host crystals regardless of mother liquor composition or viscosity. The crystallography chip has 26500 features per matrix and can easily be made with a variety of feature sizes, accommodating different crystal sizes (Mueller *et al.*, 2015). Figure 2.3 shows more detail of the chip design. The chips are loaded using a vacuum pump in humidity tents.

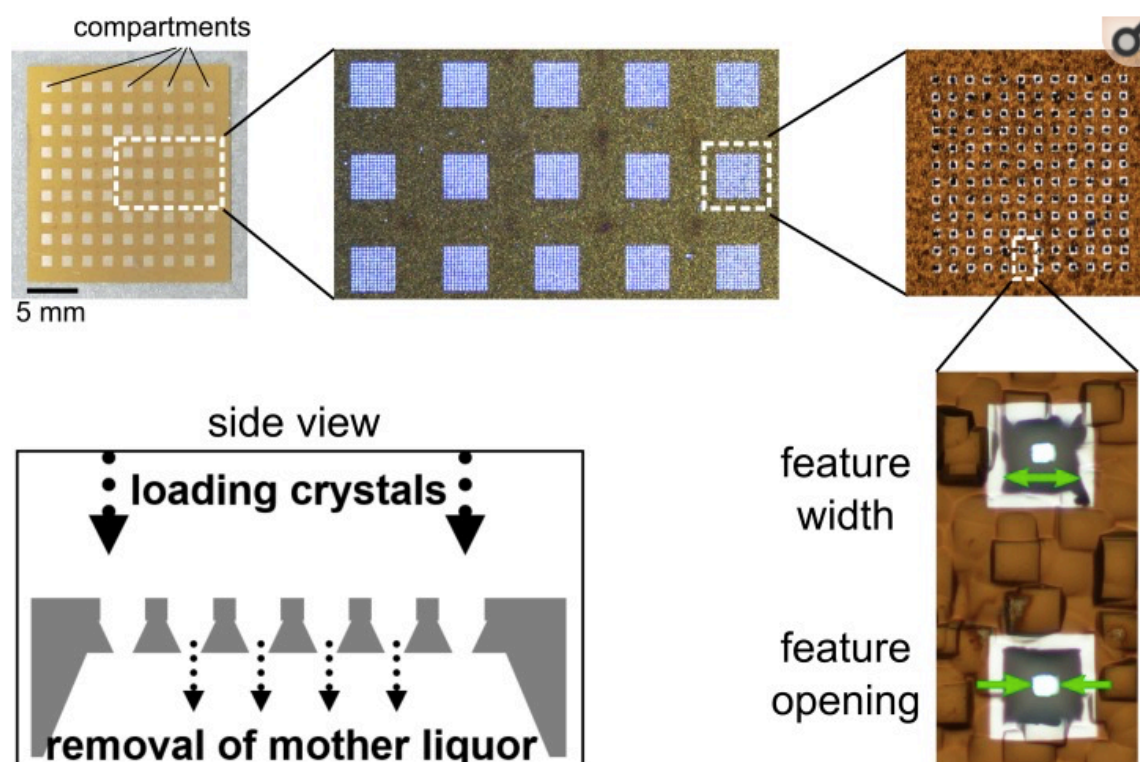


Figure 2.3: Crystallography chip design. Upper left shows a 20x20mm chip, shown in higher magnification. The 'side view' shows how the mother liquor is removed using a vacuum pump (Mueller et al 2015)

2.2.3 'Diffraction before Destruction'

Extremely short pulses of X-rays at tens of femtoseconds in duration generated by free-electron lasers can overcome the issue of dose limitations due to radiation damage. These short pulses of X-rays are shorter than the time it takes for atoms to move (Neutze *et al.*, 2000). This is known as 'diffraction before destruction', where the sample is completely vaporized by the intense pulse after giving a diffraction pattern. Doses that are thousands of times higher than conventional limits have been achieved. This means only a single flash diffraction pattern, without rotation or oscillation, can be recorded, giving rise to the 'serial crystallography' concept. (Chapman, Caleman and Timneanu, 2014).

2.2.4 Room-Temperature Serial Crystallography

With the emergence of room-temperature x-ray crystallographic experiments, microfocus synchrotrons have been developed to measure crystals *in situ*. The development of XFELs with intense pulses of femtoseconds in duration means that collecting data from room-temperature crystals can be extremely rapid, and means collecting data before the rapid onset of disorder, removing all of the issues of room-temperature data collection and radiation damage (Boutet *et al.*, 2012) that come with much higher doses.

Serial crystallography with non-cryocooled crystals also has technical advantages. Firstly, the crystals can be smaller, increasing the likelihood of growing appropriate crystals. Secondly, it avoids an increase in crystal mosaicity which can happen during cryoprotection and eliminates the need for a cryoprotectant (Guha *et al.*, 2012). The use of smaller crystals also may be easier with soaking experiments.

Chapter 3 - Methods and Materials

3.1 Bacterial transformation and DNA extraction

This project was started using an existing expression plasmid (pET16b) encoding tagless DHP-B provided by Dr Tadeo Moreno-Chicano at the University of Essex. Heat shock transformation was used to introduce the plasmid into competent BL2-DE3 strain *Escherichia coli* cells kept at -80°C in 100µl aliquots until use. A standard heat-shock transformation protocol was performed and is outlined in Table 1. Once this was performed, the cells were spread on LB agar plates supplemented with 100µg/ml of Amp and incubated overnight at 37°C, with colonies appearing the next day. Single colonies were picked and grown in 10ml of LB with Amp at 220rpm for a minimum of 18 hours. The plasmid Genejet Miniprep Kit (Thermo Scientific) was used to extract plasmid DNA. The purified plasmids were stored at -20°C until needed for future experiments.

Table 3.1: Standard heat shock protocol

Transformation Stage	Conditions
First incubation on ice	30 minutes
Heat shock at 42°C	45 seconds
Second incubation on ice	2 minutes
Addition of LB media	750µl
Incubation at 37°C	1 hour

3.2 Recombinant Protein Expression

Cultured cells from 10ml of LB agar were scaled up to 250ml flasks containing 100ml of LB media with Amp at 220rpm at 37°C. These cultures were used to inoculate large 2 litre flasks with 1 litre of LB, 100µg/ml Amp, 0.1mM hemin chloride and 0.1mM iron citrate, grown at 37°C at 220 rpm in Innova-43 incubator shakers. When $OD_{600} \approx 0.8$, assessed with a Cary60 (Agilent) UV-vis spectrophotometer, the flasks were induced with 0.5mM IPTG and the conditions changed to 25°C and 200rpm.

These were left overnight or a minimum of 18 hours.

After, the cells were pelleted at 4000rpm for 20 minutes using a Sorvall Evolution RC Centrifuge. The pellets were then resuspended in 100mM potassium phosphate buffer (Kpi) at pH8.

3.3 Protein Purification

The resuspended cell pellet was then lysed using an EmulsiFlex-C5 cell disrupter (Avestin), passing through three times to ensure complete breakage of the cell membrane. A salt cut was then performed using 50% ammonium sulphate. This was centrifuged at 18000 rpm for 20 minutes in the Sorvall Evolution RC Centrifuge. The precipitation recovered by this centrifugation was resuspended in 100mM Kpi buffer with 2M ammonium sulphate. Hydrophobic interaction chromatography was then performed (phenyl sepharose column, Pharmacia Biotech) with a salt gradient using an AKTA purification platform. Fractions were collected and pooled based on purity determined using UV-Vis and SDS-PAGE. This was followed by size exclusion chromatography (G75 column, GE-healthcare) again using an AKTA purification platform. As the protein was present in both oxyferrous and ferric states, a UV-vis was performed to determine the states of samples.

3.4 UV-vis Spectroscopic Analysis

A Cary-60 UV-vis spectrophotometer (Agilent) was used to determine the oxidation state of the samples as well as the purity. These experiments were carried out at RT, with a wavelength of 200-800nm, using a quartz cuvette (Hellma-Analytica) with 1cm light-path.

Protein concentration was determined using the Beer-Lambert equation (Eq.1)

$$A = \varepsilon_{\lambda} \cdot l \cdot c$$

where A is the absorbance of the sample, ε_{λ} the extinction coefficient at a specific

wavelength, l is the length of the light-path (fixed to 1cm) and c is the concentration of the protein sample. Extinction coefficients at the ferric Soret ϵ_{406} obtained from literature (de Serrano et al 2007). The purity of the protein was calculated from UV-vis spectroscopy by obtaining a value for each sample of the $R_z = A_{406}/A_{280}$ ratio. The protein was then concentrated down using Vivaspin-10 devices (Sartorius) until the desired concentration (around 15mg/ml for single crystal experiments and 30mg/ml for microcrystals) was obtained.

3.5 SDS-Page Gel Electrophoresis

SDS-Page electrophoresis was performed with 15% acrylamide gels after each step to check for protein purify and for the selection of fractions from the AKTA purification platform for pooling. Gels were run at 120V for 90mins on a Bio-Rad electrophoretic system. These gels were stained with Coomassie blue dye for 20-30 minutes, then destained with a solution containing: 30% methanol, 10% glacial acetic acid and 60% RO water. The destain was changed every 30 minutes until bands were visible, before being inspected on a lightbox.

3.6 Crystallization

DHP-B was crystallized according to previously established conditions (Zhang *et al.*, 1996) using the hanging-drop vapour diffusion method with protein concentrations of between 12mg/ml and 15mg/ml. The protein sample was in 20mM MES buffer at pH6.5. Crystals were manually optimised in 24-well crystallization plates with a hanging-drop volume of 4-6 μ l with a 1:1 or 1:2 volume ratio protein to reservoir, with the reservoir containing: 12-20% PEG 4000 and 150-250 mM ammonium sulphate. The plates were left at 4°C in the crystallisation fridge, and crystals appeared after 2-3 days. Some crystals were cryoprotected by soaking for 10-20 seconds in 20% glycerol before flash-freezing in liquid nitrogen. Crystals were fished and mounted on micro-loops (MiTeGen) and then mounted in pucks and placed in a

pre-cooled dewar, to ensure a stable temperature while the dewar is transported.

3.7 Crystal Soaking

DHP-B crystals were soaked to produce different complexes. The crystals were soaked in prepared crystallisation solution (20% PEG-4000 and 20mM ammonium sulphate) with 5-10mM of ligand, such as 7-bromoindole and 2,4-dichlorophenol (DCP). The crystals were soaked in ligand for 5-10 minutes and cryocooled by liquid nitrogen after soaking in glycerol as the cryoprotectant for 10-20 seconds.

3.8 Single-Crystal Measurements

3.8.1 Conventional Synchrotron Data collection

Single crystallography data was collected at Diamond Light Source and Swiss Light Source. Data was collected and cryogenic temperatures (100K), with X-ray wavelengths between 0.8-1Å. Experimental parameters, such as exposure time, % transmission and detector distance, were chosen for each experiment and controlled using the software already present at each synchrotron beamline. Diffraction spots were indexed and integrated using the programme XDS (Kabsch 2010) at Swiss Light Source and using the processing pipelines available at Diamond Light Source. Using XDS, the input file XDS.INO was modified for each crystal. This file contained important experimental and sample information, such as detector parameters, space group, resolution range, rotation and wavelength. The output files from XDS were XDS_ASCII.HKL and were used as input files for *Aimless*, which assess the space group probability and scale and merge the diffraction data, producing an mtz file. *Aimless* also produces a table of data quality statistics such as the resolution, $CC_{1/2}$ and R_{merge} . The mtz file was then used for molecular replacement and refinement.

3.9 Serial crystallography methods

3.9.1. Batch crystallization

For fixed-target serial crystallography, large amounts of microcrystals were needed, and so was produced using batch crystallisation according to established procedure

(Tadeo Moreno-Chicano *et al.*, 2019). Batch crystallisation was performed by growing crystals directing in solution, mixing protein and crystallization conditions. Batches were set up in 1.5ml Eppendorf tubes. Microcrystals of 15-35 μ m in size were grown in batches of 400-500 μ l. These batches were grown at 4°C in the crystallization fridge, and crystals would appear within 5-8 days, with the first crystals appearing after one day. The batch conditions are outlined in Table 3.2. Batches were kept on ice, and vortexed for ~30 seconds.

Table 3.2: Batch conditions for the growth of microcrystals of DHP-B

Component	Condition
Protein Concentration	~30mg/ml
Buffer	20mM MES pH 6.5
Protein-condition ratio	1:4 or 1:3
PEG-4000 %	37%-28%
Ammonium Sulphate concentration	200mM

3.9.2. Fixed target crystallography

For work carried out at Diamond (SSX) and SACLA (SFX), the ‘Diamond chip’ was used, developed at Diamond Light Source (Owen *et al.*, 2017). This is shown in Figure 3.1, where a silicon chip is held in a frame holder and sealed with mylar sheets. The diamond chip contains 25,000+ apertures, with variable sizes depending on the size of the microcrystals being looked at. These chips were specifically designed to provide a support for the microcrystals while also keeping background to a minimum during diffraction experiments.

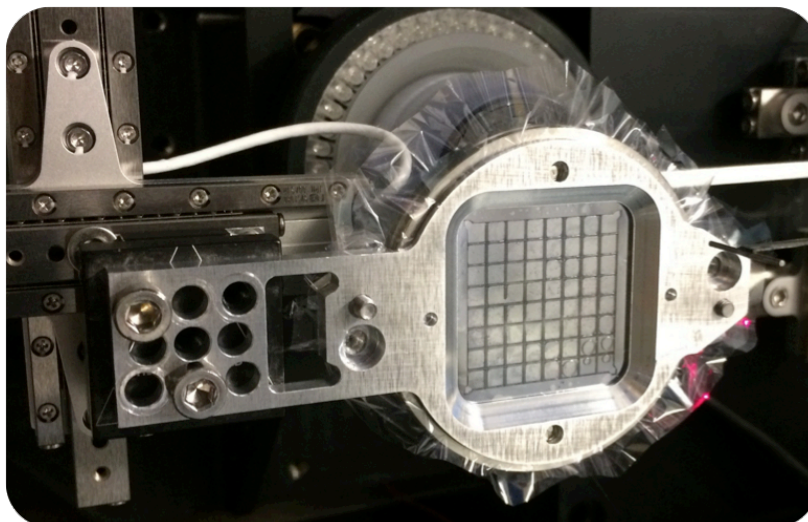


Figure 3.1 the 'Diamond Chip' used at Diamond Light Source at beamline i24 for SSX crystallography. This chip is able to hold >25,000 crystals, and is mounted on a specific 'xyz' stage.

Chips were loaded inside a humidity 'hood', with a humidity level of 70%-80%. Chips were glow-discharged (Pelco-easyGlow), and a microcrystal solution of 150-200 μ l loaded onto the chip. The excess liquid was removed using a vacuum pump, and the chip was then sealed in the frame holders with mylar.

3.9.3. Chip Screening

Data was collected at Diamond Light Source by moving through the surface of the chip, with an automated chip screening routine to ensure fresh microcrystals were being hit each time. The chips were mounted and aligned with the help of three fiducials in each corner of the chip. Each of the 25,600 positions hit at a frequency of 30 Hz. Each chip took approximately 14 minutes, unless only the central part of the chip was screened (due to lack of crystals or a test run), where screening took half of the time.

3.9.4. SFX Data Processing

During the beamtime at SACLA, the XFEL data was preliminarily indexed and on-the-fly estimations of hit rates were done using the program *Cheetah* (Barty *et al.*, 2019). *Cheetah* also selected the frames with protein diffraction, making the volume of data smaller.

SFX data were obtained as *h5* files, and processed using *CrystFEL* (White *et al.*, 2012). *CrystFEL* is a software suite containing programs dealing with viewing, indexing, integrating, merging and evaluating the quality of the data. All of these programs are run from the command line. The usual workflow is indexing then scaling and merging.

3.9.4.1 Indexing

The first step is indexing, which was carried out using *indexamajig*, part of the *CrystFEL* program suite. The parameters used for this was tested with the ‘test’ protein, with known cell parameters and consistent diffraction quality. The commandline for *indexamajig* is and the parameters with their definitions are shown in Table 3.3

```
$ indexamajig -i files.lst --peaks=peakfinder8 -
threshold=300 min-gradient=90000 --min-snr=5 --int-
radius=3,4,7 --indexing asdf -g detector geom -p
parameters.cell -o testingrun.out -j 20
```

Command	Definition
-I files.lst	Path to the input diffraction data
--peaks=peakfinder9	Peak-finding algorithm <i>peakfinder8</i>
-threshold=300	Threshold of photon counts for detecting diffraction spots
--min-gradient=9000	Minimum square gradient for detecting peaks
--int-radius=3,4,7	Sets the inner, middle and outer radii for three ring integration
--indexing=asdf	Indexing using the asdf <i>CrystFEL</i> algorithm
--min-snr=5	Minimum value for signal to noise ratio

-g detector.geom	File containing detector geometry
-p parameters.cell	Symmetry information for microcrystals
-o.testrun.out	Generates output file
-j=20	Number of processors with a max value of 20

Each chip initially processed this way, creating a 'stream.out' file containing indexed reflections and the parameters examined using the program *cell_explorer*, generating distribution plots for each parameter.

```
$ cell_explorer firstRun.out
```

A second run of *indexamajig* was then run with the optimized cell parameters, generating a refined output file 'secondRun.out'. The detector geometry was also refined using *detector-shift*.

```
$ detector-shift .out secondRun.out detector.geom
```

detector shift also plots the disagreement, shown as detector shift in x and y, between the location of experimental and predicted spots in every indexed frame. A refined geometry file was then produced named 'refined.geom'. The final file 'finalRun.out' for a different chip with the sample could then be joined to form a new file for scaling and merging. The command used to combine the output files:

```
$ cat chip1.out chip2.out chip3.out > allchips.out
```

3.9.4.2 Merging

Merging was carried out using the program *process_hkl* in the *CrystFEL* suite on the recommendation of Dr. Takanori Nakane. *Process_hkl* is included in the script

merge.sh, which can be run with the reflections from *indexamajig* as the input file.

3.9.4.3 Data Quality Statistics

Data quality statistics (CC and R_{split}) were generated using *CrystFEL* script *stat.sh*, using the output from merging (hkl files). Symmetry information was also needed as input for *stat.sh*.

R_{split} is used exclusively for serial crystallography to estimate the quality of the diffraction data. Data is firstly split into two halves and then merged independently and compared. It is normally shown as a percentage, with the closer to 0 the better. Estimate guidelines for R_{split} are used: a value of 35% is acceptable, below 20% is better and below 10% is excellent.

Correlation coefficient (CC) is equivalent to the $CC_{1/2}$ statistic used in conventional X-ray crystallography (Diederichs and Karplus, 2013). CC compares the similarity between the data when split into two halves, the values given are between 1 and 0 with 1 being identical and 0 being no similarities at all. A $CC > 0.5$ is usually used to make the resolution limit.

3.9.4.4 Producing an mtz file for molecular replacement and refinement

An mtz file is produced using the *CrystFEL* script 'create-mtz'. The hkl file from the merging processes was used as input, as well as the cell parameters for the crystal. The mtz file was then used with the CCP4i2 suite for molecular replacement and ultimately refinement of the protein structure.

3.9.5. SSX Data Processing

SSX data processing was developed at Diamond, with chip aperture and data collection parameters (detector distance, exposure time) tailored to the sample (Owen *et al.*, 2017) Serial crystallography data was measured at Diamond Light Source at beamline I24, and stored in *cbf* files, processed using *dials.stills_process* in the *DIALS* programme (Winter *et al* 2018), carrying out indexing and integrating diffraction spots. *dials.stills_process* was developed specifically for serial data

processing, and so no rotation of the crystals are involved like typical synchrotron data collection. The integrated data is put into *PRIME*, where it is scaled and merged (Uervirojnangkoorn *et al.*, 2015) generating an mtz file. This mtz file can then be used for molecular replacement and refinement. Mtz files were also used to assess the quality of the data.

3.10 X-ray data

X-ray data was collected at Diamond Light Source (DLS) and Swiss Light Source (SLS). All x-ray data was collected at 100K and at an X-ray wavelength of 0.8Å. Experimental parameters were decided based on each sample and were controlled using the in-house software at each synchrotron beamline. The mtz file produced was used for molecular replacement and refinement.

3.11 Molecular Replacement and Refinement

As the structure of DHP-B has been previously determined, the structures were solved using molecular replacement using *Phaser* within either Ccp4i2 suite or Phenix. For SSX data, the mtz with the experimental data were inputted into *Phaser* with the pdb file obtained from the Protein Data Bank (www.pdb.org). For serial data, an existing model was provided for DHP at room temperature (Hough and Moreno-Chicano 2019). The success of molecular replacement and refinement was assessed using the R-value generated, aiming for the R-free and R-work to be close together and below 0.2, as well as observing the structure in Coot. Initially, in *Phenix*, the mtz and pdb files were analysed using *phenix.xtriage* (Zwart, Grozse-Kunstleve and Adams, 2005) to check for potential twinning and overall completeness of the data.

For each structure, multiple rounds of refinement and model building were carried out. Refinement was carried out using *phenix.refine* in Phenix (Afonine *et al.*, 2012) or Refmac5 (Murshudov *et al.*, 2011) in the ccp4i2 suite. Manual refinement would also be carried out using *Coot* (Emsley *et al.*, 2010) after automated refinement produced an electron density map and a difference map. Manual rebuilding used

both these maps to indicate any differences in structures, such as conformational inaccuracies. The contour map was visualised at a 1σ contour level. After the initial refinement, both 'cartesian' and 'torsion angle' simulated annealing was performed to remove model bias, ensuring we have the correction structure. Manual refinement was used to check that this had worked, and had not displaced important features i.e. the haem centres and the general protein backbone and amino acid sidechains. At this point, water molecules and ions were added. Ligands were modelled in at the very end of refinement. Rounds of model refinement were then performed until the model could no longer be improved.

Chapter 4 – Single Crystal DHP Crystallography

4.1 Introduction

As discussed previously, DHP is a multi-functional haem protein of *A.ornata* that *in vivo* oxidises trihalophenols produced by competing organisms in its ecosystem. DHP has its main function as a peroxidase, but can also carry out many other additional functions, such as an oxidase and a peroxygenase (Malewschik *et al.*, 2019). DHP's main substrates are halogenated compounds. The aim of this study was to compare structural differences between DHP-B bound to various ligands to determine the functionality of the protein.

4.2 Protein Purification

The first step of protein purification was a salt fractionation, adding ammonium sulphate up to 50% (precipitation point for DHP). After centrifuging at 18,000rpm for 20 minutes to remove the precipitate. This was resuspended in 20mM KPi buffer at pH8 with 2M ammonium sulphate. The protein was loaded onto a hydrophobic interaction chromatography phenyl-sepharose column on an AKTA-purifier, running a salt gradient from 0 to 2M ammonium sulphate. If two bands appeared, both ferric and ferrous DHP is present in the sample (Figure 4.1) and the bands were collected separately. UV-vis absorbance spectroscopy was performed on all fractions, with all pure fractions pooled. This was then run through a G-75 size exclusion column, with all fractions again assessed by UV-vis and all pure fractions pooled. The pure protein samples were concentrated using a Viva-spin-10 concentrator with a 10kDa cut-off, and the buffer was exchanged to the final 20mM MES buffer at pH 6.5. The protein was concentrated until a ~15mg/ml concentration was achieved, with the concentration calculated by UV-vis spectroscopy using the extinction coefficient at haem absorption maximum DHP-B: $\epsilon_{406} = 116.4\text{mM}^{-1}\text{cm}^{-1}$ (D'Antonio *et al.*, 2010)

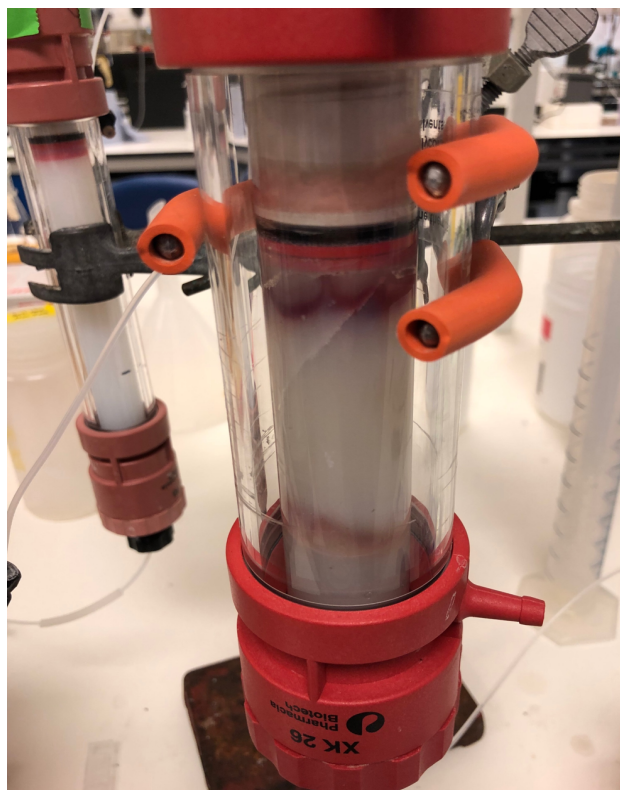


Figure 4.1: Phenyl-sepharose column with both ferric (top) and ferrous (bottom) bands bound during HIC purification. The column was then run on an AKTA purifier using a salt gradient, and both bands collected separately.

4.3. Crystallisation

DHP-B crystals were grown using hanging-drop vapour diffusion method at 4°C using the crystallization conditions outlined in literature (de Serrano *et al.*, 2007). A protein concentration of 8-15mg/ml was used and drops of 6µl with a 1:1 volume ratio protein to reservoir. These were equilibrated in a 24-well crystallization plates against a reservoir containing 20-33% PEG 4000 and 150-250mM ammonium sulphate. Crystals would appear after 4 days. The DHP crystals would grow to 50-400µm.

4.4. Crystal Soaking

DHP-B crystals were soaked in crystallization solutions with 20% PEG and 200mM ammonium sulphate with added 5-20mM concentrations of ligand. Ligands used in this work include: 7-bromoindole, 5-bromoindole, 4-bromophenol and 2,4-

dichloropheno (DCP). 5-bromo indole was selected as the ligand appears to bind in the ferric structure directly above the haem (tested by Ghiladi group) All are able to interact with DHP as it's main function as a peroxidase means it performs oxidation of trihalophenols to dihaloquinones, and its ability to accommodate large aromatic ligands. For DCP and 7-bromoindole, stock solutions of 0.5M were prepared in water. For 5-bromoindole and 4-bromophenol, stock solutions were prepared in ethanol, keeping levels below 2% ethanol to avoid damaging the crystals. Crystals were left in the soaking solutions for 5-10 minutes before being transferred into a cryoprotectant solution of reservoir solution with 20% glycerol added, before being flash cooled in liquid nitrogen.

4.4.1 Ligand Bound Structure of DHP-B with 7-bromoindole

The first single crystal structure collected to test the soaking conditions was using the ligand 7-bromoindole. Ferric DHP crystals were soaked in crystallization condition with 5mM of 7-bromoindole for 5-10 minutes before soaking in the cryoprotectant (20% glycerol) and flash-cooled in liquid nitrogen. The crystal structure was measured at Diamond Light Source (DLS). The diffraction data was processed as outlined in Chapter 3, and the data collection and refinement statistics are shown in Table 4.1. The overall structure is shown in Figure 4.2

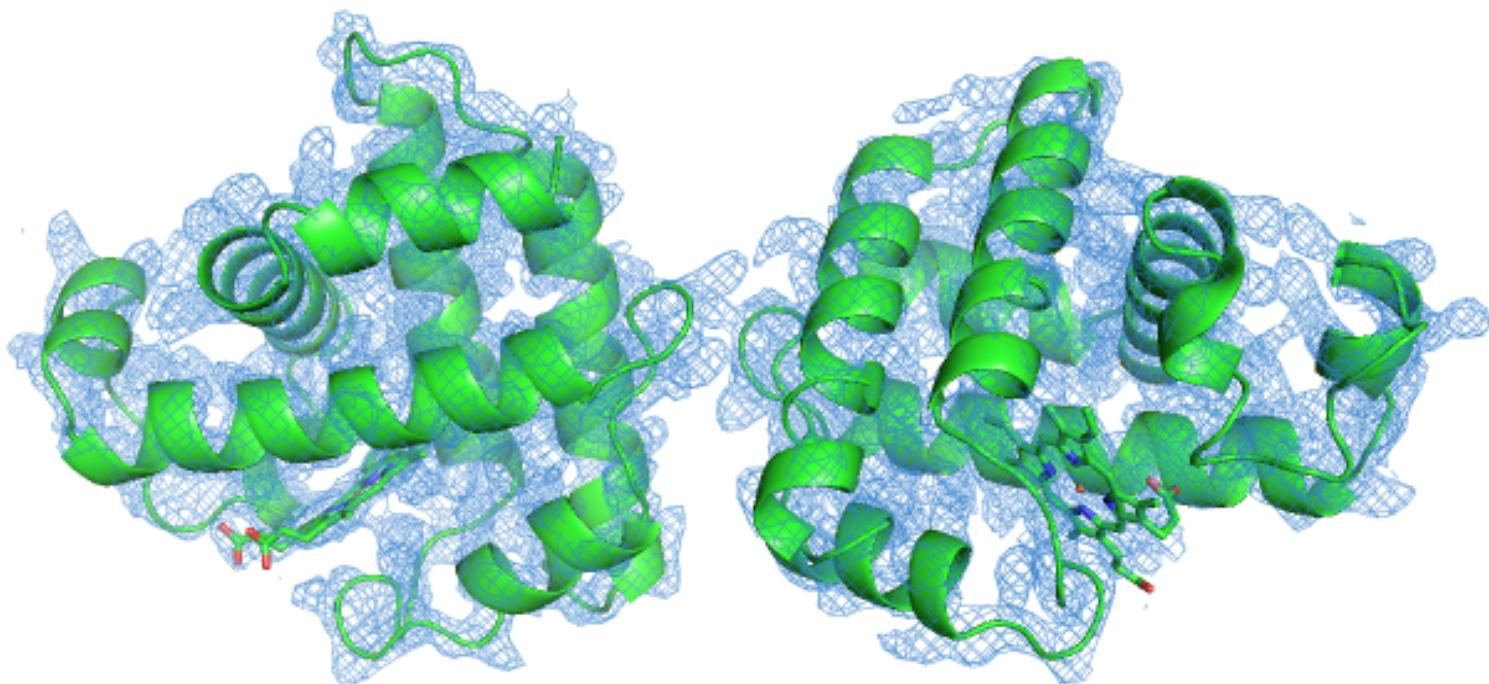


Figure 4.2: overall structure of DHP-B structure measured at DLS. Electron density shown as blue mesh.

The ligand was not present in either haem pockets, so for further soaking experiments we used a higher concentration of ligand in the soaking conditions. The structures of both haem A and B are shown in Figure 4.3. Note in Figure 4.3 that the distal His (His55) is in a 'closed' conformation, inside the distal pocket. The His55 and haem Fe distance at haem A and B is 2.60 Å and 4.09 Å respectively, implying that the His55 at haem A is in the 'closed' position, while being 'open' at haem B.

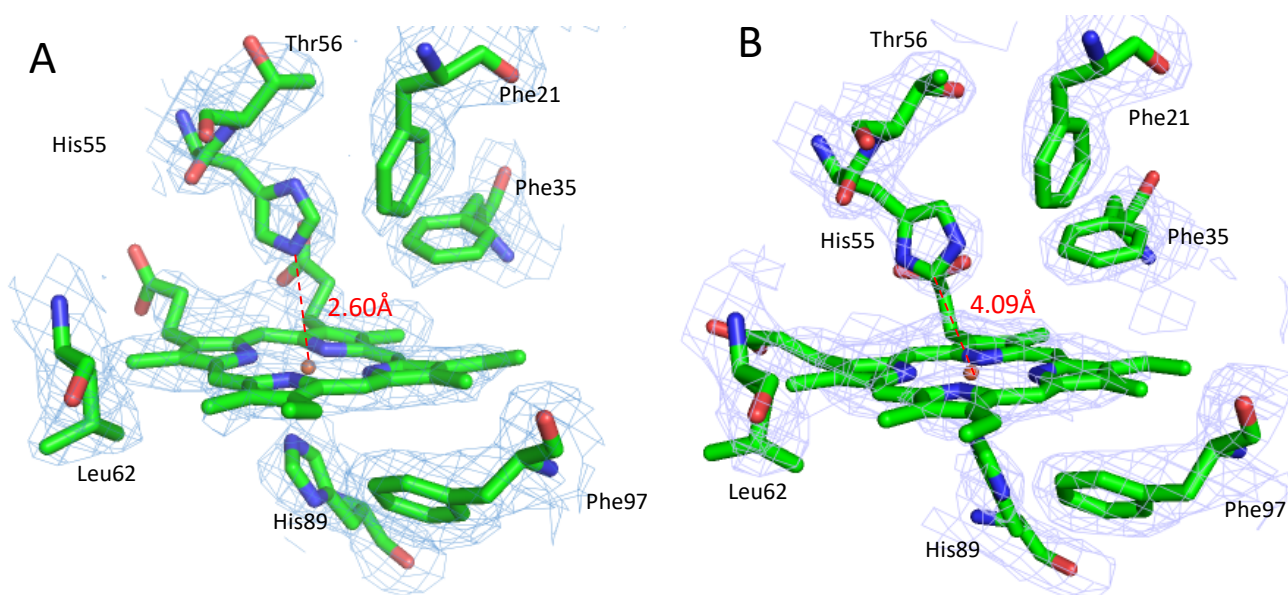


Figure 4.3: Structure of Haem A and Haem B from the structure of DHP crystals. Electron density is shown as blue mesh. Contoured to 1σ . Measured at beamline i03 at Diamond Light Source (DLS; Didcot, Oxford)

Table 4.2: Refinement statistics and data collection for DHP-B 7-bromoindole complex measured at DLS. Values in brackets represent the stats from the highest shell of resolution

	7-bromoindole
Space group	P2 ₁ 2 ₁ 2 ₁
Unit cell (Å)	A = 61.01 b= 67.86 c=68.00 $\alpha=\beta=\gamma=90^\circ$
Resolution (Å)	48.04 - 2.36
Wilson B-factor (Å ²)	43.9
Unique reflections	11728
I/ σ (I)	7.7
CC _{1/2}	1.0 (0.803)
Completeness (%)	96.84 (99.5)
R _{work}	0.19 (0.27)
R _{free}	0.29 (0.36)
Reflections for R-free	11725 (1190)
Rmsd bond lengths (Å)	0.014 (0.002)
Rmsd bond angles (°)	1.62 (0.36)
Ramachandran	
favoured	92.22%
Clashscore	10.33
Average B-factor	55.14

4.5 DHP-B Crystal Soaking Experiments at SLS

The crystals were prepared for DHP-B and 4-bromophenol and 7-bromophenol as prepared for collection at Swiss Light Source (SLS) at beamline PXII. Originally, the plan for this was the collect more ligand soaks, and to produce a ‘before and after’ UV-vis spectrum and Resonance Raman. There was also the possibility of a dose series. The beam at SLS was 100x100 microns. The beam was set to 1% Flux before changing to 10% flux to try to obtain any diffraction. We were unable to measure Resonance Raman spectrums as the crystals were too small. Due to a shipping error,

all usable crystals were too icy to obtain a diffraction pattern, with too many ice rings.

4.6 Comparisons from Literature

This section includes comparing some of the DHP ligand complexes from literature. Firstly, the structure of DHP-B with ligand 4-nitrophenol (PDB: 5CHQ), also known as PNP, is a phenol used as a fungicide and is considered an organic pollutant. The structure of 4-nitrophenol is shown in figure 4.3A. 4-Nitrophenol binds closely to the haem cofactor. The ligand is found in both active sites of the dimer in the asymmetric unit. His55 (distal His) swings out of the distal pocket to allow 4-nitrophenol to bind. This is shown in figure 4.3A (McCombs *et al.*, 2016)

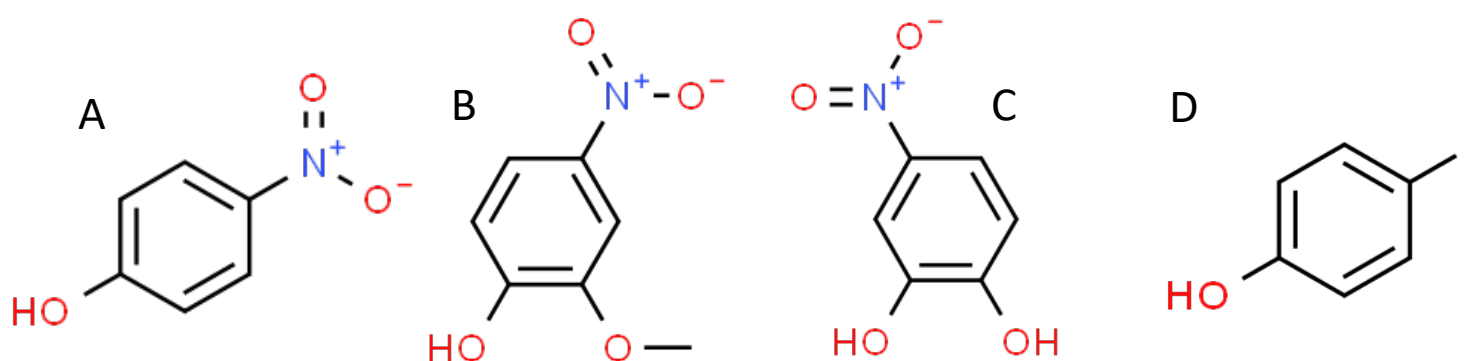


Figure 4.3: Structures of the DHP-B ligands. A: 4-nitrophenol B: 4-guaiacol C: p-nitrocatechol D: 4-iodophenol

The ligand 4-Nitroguaiacol (PDB: 6CH5) is also analysed, chosen due to its presence in the environment that *A.ornata* inhabits, its structure is also shown in Figure 4.3B. The 4-nitrophenol binds only at one active site (haem A), while at haem B we see a hemichrome species with His55 interacting directly with the iron, forming a hexacoordinated complex (McGuire *et al* 2018). At haem A, the ligand is bound in the distal pocket but doesn't form a complex with the haem iron, due to binding higher in the pocket and instead forms hydrogen bonds with the aromatic ring of Phe60 (Figure 4.3B). The ligand p-nitrocatechol, is also an organic pollutant and a phenolic derivative of PNP (PDB: 5CHR), with its structure shown in Figure 4.3C. p-nitrocatechol binds to both haem A and haem B. p-nitrocatechol binds in a binding

site in close proximity to the haem cofactor. The distal histidine His55 swings out of the distal pocket to allow binding (Figure 4.3C). At haem B, the iron in the haem is bound to a O_2 molecule present in the binding pocket. p-nitrocatechol complexes with the haem iron at haem A (McCombs et al 2016) The last ligand to compare is 4-iodophenol, with the structure being shown in Figure 4.3D. 4-iodophenol binds in the distal pocket (PDB: 1EWA), with the distal histidine swung out (Figure 4.3D). Again, the 4-iodophenol binds to a pocket close to the haem cofactor (La Count et al 2000)

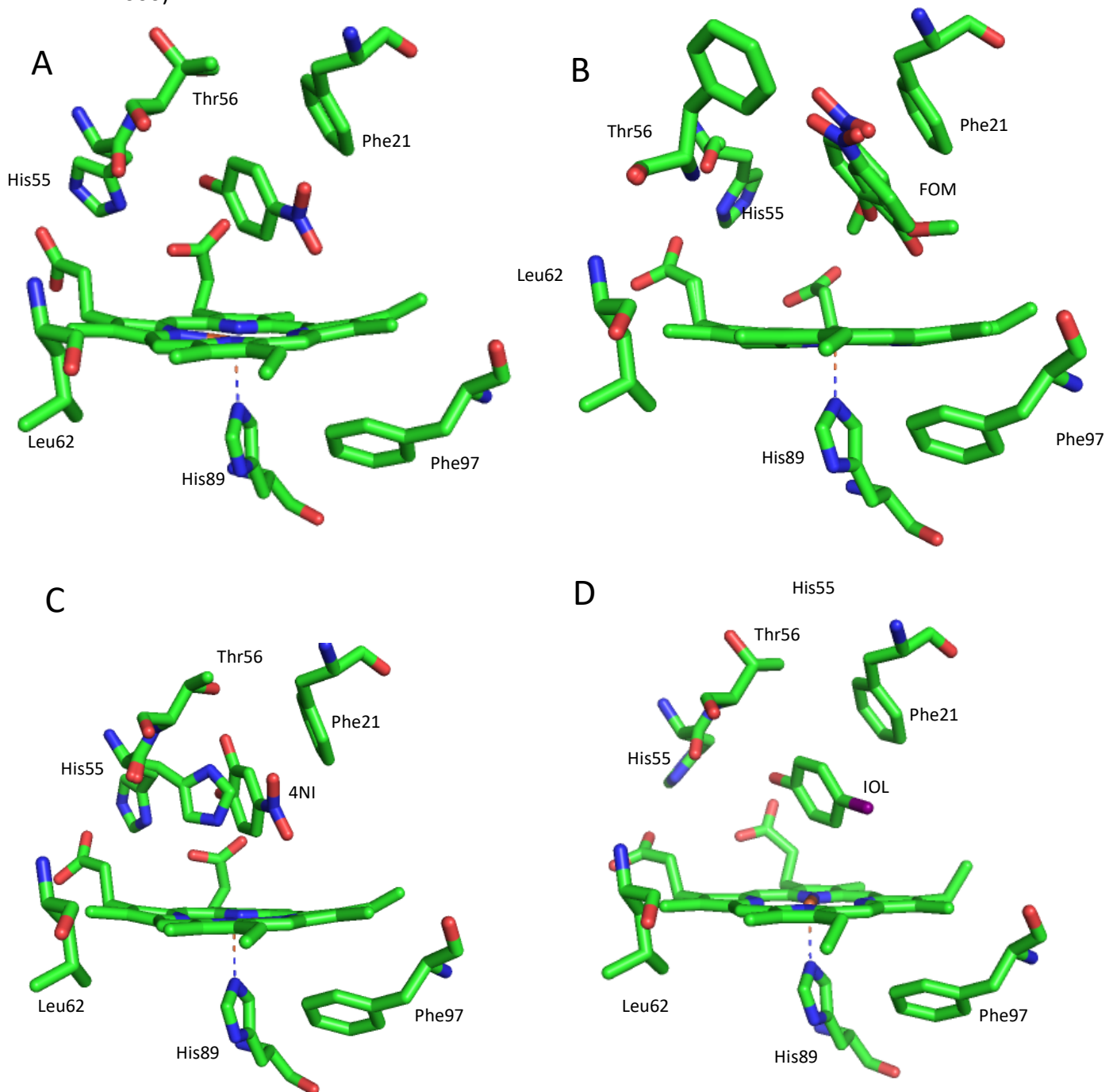


Figure 4.3: Binding modes of the distal pocket with DHP-B. The ligands used were A: 4-nitrophenol B: 4-guaiacol C: p-nitrocatechol D: 4-iodophenol. Hydrogen bonds are shown as dashed lines and the most relevant surround residues named.

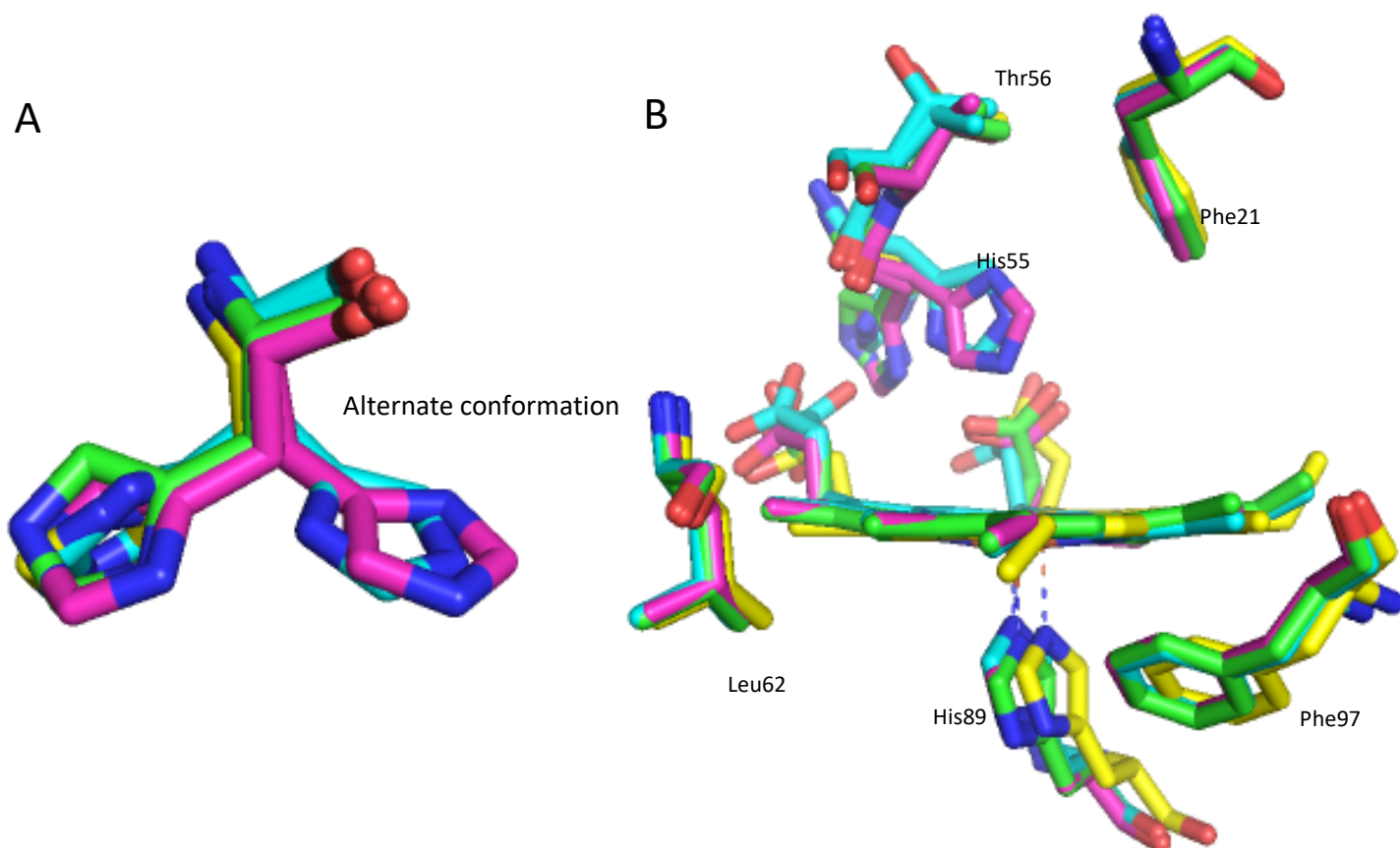


Figure 4.4: Superposition of the different binding moves around the haem (monomer A) of DHP-B ligand complexes. Ligands have been removed for clarity. (A) shows more clearly the conformational changes of His55 depending on the ligand bound. Green: 4-nitrophenol Light Blue: 4-nitroguaiacol Pink: p-nitrocatechol Yellow: 4-iodopheno

Table 4.3: Bond and ligand distances for 4-nitrophenol, 4-guaiacol, p-nitrocatechol and 4-iodophenol with DHP-B.

Complex	His55-Ligand Distance (Å)	Fe-Ligand Distance (Å)	Hydrogen Bonds (Å)		
			Water to Ligand	Ligand to Thr56	Ligand to Phe60
4-nitrophenol	3.82	4.09	-	-	-
4-guaiacol	7.55	5.22	3.1	-	1.33
p-nitrocatechol	3.16	4.18		2.57	
4-iodophenol	4.75	3.56	-	-	-

4.7. Discussion

DHP has been extensively investigated using single crystallography, with 78 structures deposited on the protein data bank (Berman *et al.*, 2002). In this chapter, the main focus is on the different binding modes of DHP-B, with four structures from the PDB, as well testing the ligand soaking conditions for DHP-B single crystals.

The structure of DHP-B soaked with 7-bromoindole was measured at beamline i03 at Diamond Light Source, UK. The structure was determined at 2.36 Å, a lower resolution to other previously solved structures, this is probably due to the age of the crystals being a few months old. The haems (Figure 4.3) show the distal histidine in a 'closed' conformation. There is no density above or below the haem plane to suggest a ligand is bound, and so the ligand soak was unsuccessful. Originally, 5mM was used as it was the lower end of the concentrations used in previous ligand soaking experiments (Tadeo Moreno-Chicano *et al.*, 2019) however due to no ligand being bound in this structure, for future soaks a higher concentration of ligand was used.

Comparisons of ligand binding from literature show how varied binding modes can be in DHP-B depending on the ligand. In all ligands, the distal histidine is 'swung out' to allow the large aromatic ligands to bind. His55 is a key residue in binding ligands, contributing to whether ligands are allowed to bind (Francesco *et al.*, 2010). P-nitrocatechol shows two conformations of His55, and has the shortest His-Fe distance and is closest to a hydrogen bonding distance, suggesting His55 may be acting as a hexacoordinate ligand. P-nitrocatechol also has a relatively short ligand-Fe distance compared to the other ligands outlined here, however too large to be considered a 'bonding distance'.

Chapter 5 - Serial DHP Crystallography

Serial crystallography using 'the Oxford chip' developed at Diamond Light Source. DHP microcrystals are mounted on silicon nitride chips (discussed in Chapter 3). Serial crystallography was performed at synchrotron and XFEL beamlines. Batch crystallography conditions have previously been characterised. Ligand soaking using microcrystals were also used, providing damage-free SFX structures of these complexes.

5.1 Batch Crystallization

Production of DHP microcrystals was previously optimized by Dr Tadeo Moreno-Chicano, and adapted by me. Previously, by Dr Moreno-Chicano, thousands of DHP microcrystals of 15-35 μ m in size in a single batch of 400-500 μ l. Microcrystals were grown at 4°C within 5-8 days. The batch conditions used during this project are outlined in Chapter 3, where the outlined conditions were not successful, with the results being unpredictable. This may be to do with the purity of each round of expression/purification of DHP-B. The trials of attempting to produce microcrystals of DHP are outlined in Table 5.1. Fresh PEG-4000 was used for each batch, as when older PEG solutions were kept, the solution would precipitate and appear cloudy. For the experiments outlined, older crystals from previous experiments were used.

Table 5.1: DHP-B screens to produce microcrystals based on previously successful production of microcrystals.

Condition	Final concentration in		Protein (mg/ml)	Set up		Outcome
	Batch			Ratio Protein:	Total Volume	
	% PEG 4000	AmSO4 (Mm)				
DHP1	36%	200	27	1 to 3	50	No crystals
DHP2	37%	200	28	1 to 3	50	No crystals
DHP3	38%	200	29	1 to 3	50	No crystals
DHP4	36%	200	30	1 to 3	50	Precipitated
DHP5	37%	200	27	1 to 4	50	No crystals
DHP6	36%	200	28	1 to 4	50	No crystals
DHP7	38%	200	30	1 to 4	50	Precipitated
DHP8	37%	200	30	1 to 3	50	No crystals
DHP9	38%	200	30	1 to 3	50	No crystals

5.2 Synchrotron Serial Crystallography (SSX) at Diamond Light Source

DHP microcrystals grown in previous projects were taken to beamline i24 at Diamond Light Source (Didcot, Oxford, UK). The microcrystals were measured using ‘Diamond chips’. ‘Diamond chips’ are made from silicon nitride and contain 25,000+ apertures (shown in Chapter 3), obtaining low-dose structures using diffraction data from thousands of crystals. The aim was to get SSX data of DHP-B ligand bound structures to compare to single crystal cryogenic structures as well as structures from our XFEL beamtime at SACLA.

Ligand soaking experiments were performed with microcrystals by making stock solutions of ligands used of 500mM and adding these directly to the Eppendorf of microcrystals to a concentration of 10-50mM.

Microcrystals were loaded onto chips in a humidity controlled hood (previously discussed in Chapter 3). One chip (Balzac) was measured at i24, to test the ligand soaking conditions didn’t dissolve the crystals and to also check that the crystals

were diffracting as expected. This chip was DHP-B soaked with 5mM DCP. The next chip (Forster) was measured for the structure of DHP-B and 7-bromoindole and were screened following the screening protocol also in Chapter 3. However, the first run of the chip was saved to the wrong folder, and so the run was stopped. We then reran the chip, excluding the column that had already been measured. The experimental parameters for these two chips are shown in Table 5.2

Table 5.2: Experimental parameters and statistics for chips loaded at i24, Diamond Light Source (Oxford, UK) at 12800 eV energy

Name of Chip	Beam Size (μm^2)	Detector Distance (mm)	Exp Time (ms)	Trans (%)	City Blocks	Measured Frames	Indexed by Dials	Hit rate (%)	Res Crystals (
Balzac	7x8	330	10	50	c1-c4	-			
Forster1	7x8	340	10	-	Full chip	25600			
Forster2	7x8	340	10	-	7 columns	22400			

5.2.1 Processing of SSX data

The data from I24 was processed using *DIALS* (Winter et al 2018) for indexing, and *PRIME* (Uervirojnangkoorn *et al.*, 2015) was used for scaling and merging. *DIALS* contains a program called ‘dials.stills_process’ which is specifically developed to index diffraction patterns from many crystals. The integrated and merged files were then put into *PRIME* where an mtz file was generated. This mtz was loaded into ccp4i2 and the usual program of *Phaser* for molecular replacement, *Refmac* for refinement and *Coot* to build the model.

‘dials.still_process’ was run from the command line and submitted to the Diamond cluster. As an example, SSX data for **7-bromoindole with DHP-B** is used as an example. To run *DIALS*, the file ‘dhp.phil’ had to be generated, containing information such as space group and cell parameters and ‘chip_name_process.sh’, the executable script to run a *DIALS* job. In Figure 5.1 shows the first portion of the

'dhp.phil' script.

```
#### uncomment below to use mask file (should be detZ/energy-specific to mask Si reflections and b
#spotfinder.lookup.mask=/dls/i24/data/2017/nt14493-63/processing/stills_process/mask_220.pickle
#integration.lookup.mask=/dls/i24/data/2017/nt14493-63/processing/stills_process/mask_220.pickle
#### option below to change from default (3) minimum spot size
#spotfinder.filter.min_spot_size=2

#### optional detector geometry which may have been produced from joint refinement (detZ specific)
#geometry {
#  detector {
#    panel {
#      fast_axis = 1,0.0001385385979190844,-5.3528031635062794e-05
#      slow_axis = 0.00013842417630211348,-0.9999977182912047,-0.0021317249194795396
#      origin = -217.63276148203622,226.5777726218615,-248.5997478704318
#    }
#  }
# }

#### edit spacegroup and cell for the protein in question
indexing {
  known_symmetry {
    space_group = P212121
    unit_cell = 60.9 67.2 68.7 90 90 90
  }

  refinement_protocol.n_macro_cycles = 1
  refinement_protocol.d_min_start=2.5
  basis_vector_combinations.max_refine=5
  stills.indexer=stills
  #### below will try real space grid search method then fftld if the former fails
  stills.method_list=real_space_grid_search fftld
  #### maximum number of lattices to search for below
  multiple_lattice_search.max_lattices=6
}

integration {
  integrator=stills
  profile.fitting=False

  background {
    simple {
      outlier {
        algorithm = null
      }
    }
  }
}
```

Figure 5.1. First part of the file 'dhp.phil' used for indexing with DIALS of the chip 'Forster'.

The command 'dials.stills_process' was then used using:

```
dials.still_process ../forster/forster0007_000{00..99}.cbf
../forster/process.phil mp.nproc=20 > dhptest.out
```

where:

- ../forster/forster0007_000 specifies the path to the image files
- nproc=20 specifies the number of processors
- > dhptest.out is the output file

Once run, indexed and integrated files would appear in the working folder until the run had finished.

5.2.2 SSX Structure of DHP-B in Complex with 7-bromoindole

DHP microcrystals (grown before I took over the project) were taken to beamline i24 at Diamond Light Source, Oxford, UK. At this beamtime, the diffraction of the crystals and the hit rate were optimized. The measurement at Diamond would help to prepare for the XFEL trip to SACLA, as we would be using the same system there.

A stock solution of 500mM 7-bromoindole was added to the batch of microcrystals (measuring from 30-50 microns) to give a final solution concentration of 10mM 7-bromoindole. These microcrystals were left for 15 minutes before being pipetted onto a silicon-nitride 'diamond chip' in a humidity tent (shown in Chapter 3). The experimental set up is shown in Table 5.2

The spots for the dataset 'forster' were not strong enough to index, and therefore I was unable to continue processing the data.

5.3 Serial Femtosecond Crystallography (SFX) at SACLA

DHP microcrystals were taken to SACLA (Japan) at the BL2/EH3 beamline. Samples were transported in a flask to keep them at 4°C during the travel to SACLA. The data for oxyferrous and DHP and peracetic acid were collected on a previous trip before I started the project, and an Eppendorf of DHP microcrystals suspended in their crystallisation solution were brought to SACLA and were measured. After I took over the project in 2019, one chip of DHP soaked with 7-bromoindole was loaded onto silicon nitride chips ((Mueller *et al.*, 2015) and mounted in the XFEL hutch. The chip names and what they contained as well as the years measure are shown in Table 5.3

Table 5.3: Chip names, their conditions and the year they were measured at SACLA, Japan. 2018 chips were measured previously by Dr Tadeo Moreno-Chicano. (* measured during this project)

Chip Name	Condition	Year Measured
Endive	DHP-PA	2018
Fostis	DHP-PA	2018
Ginger	DHP-PA	2018
Almond	DHP Oxyferrous	2018
Cockle	DHP Oxyferrous	2018
Zinger	DHP Oxyferrous	2018
Neeson*	DHP 7-Br*	2019*

5.3.1 Measured SFX Data for DHP -7-bromoindole Complex

Firstly, some pipetting was needed to detach some of the crystals from the Eppendorf tube walls as well as resuspending crystals that had collected at the bottom. For the ligand soak, 7-bromoindole was made in a 500mM stock, and 20 μ l added to 1ml Eppendorf of microcrystals. Once added, the batches were mixed and left for 5 minutes to soak before being loaded onto the chip. One chip was loaded for the ligand soaks, due to only a small amount of crystals being available. Previously, and for the oxyferrous data, four chips were needed to obtain the structures. For the soak, a six-letter name based on an actor/actress was given to the chip. For the parameters used, see Table 5.4.

Table 5.4 SFX parameters used at SACLA beamline BL2 EH3 for all chips measured in 2018

Component	Condition
Energy	10 KeV
Repetition Rate	30 Hz
Beamsize	2x2 μ m
FEL pulse length	10 fs pulse

Table 5.5 SFX parameters used at SACLA beamline BL2 EH3 for all chips measured in 2019

Component	Condition
Energy	11 KeV
Repetition Rate	30 Hz
Beamsize	2x2 μm
FEL pulse length	10 fs pulse

5.3.2 Processing SFX Data

Hit finding and estimation of hit rates was carried out using *Cheetah* (Barty et al 2014), producing *H5 files*. The software CrystFEL was used to process all data from SACLA, a program specifically developed for processing XFEL data. As previously stated in Chapter 3, CrystFEL contains multiple different programs run on the command line, following the usual structure of indexing, scaling and then merging. All diffraction data was indexed using *indexamajig* using:

```
$ indexamajig -i files.lst --peaks=peakfinder8 --threshold 300 -
min-gradient=90000 --min-snr=5 --int-radius=3,4,7 --indexing=asdf
-g new_geometry-predrifine.geom -p DHP.cell -o DHP.out -j 20
```

These parameters are defined in Chapter 3, and every chip was processed in this way. A ‘stream.out’ file was produced containing the indexed reflections. To demonstrate processing of SFX data, outlined is the processing of the data for **DHP in complex with Peracetic Acid**, measured by Dr Tadeo Moreno-Chicano during the 2018 SACLA beamtime. The *DHP.cell* file containing the cell parameters and symmetry for DHP, which was made using the *CrystFEL* template, including the RT parameters obtained from a previous SSX beamtime on i24 (DLS, Oxford). This .cell file is defined:

```
CrystFEL unit cell file version 1.0

lattice_type = orthorhombic
```

```

centering = P
a = 61.36 Å
b = 68.01 Å
c = 102.33 Å
α = 90 deg
β = 90 deg
γ = 90 deg

```

An initial indexing run (around 200 crystals) was done to find the parameters for the specific microcrystals. After this, the cell parameters were examined using *cell_explorer* from the CrystFEL package. This generated histograms for each cell parameter, as well as calculating the mean value and standard deviation (Figure 5.2). The command to run this is shown here for the chip ‘Fostis’

```
$ cell_explorer firstRun_fostis.out
```

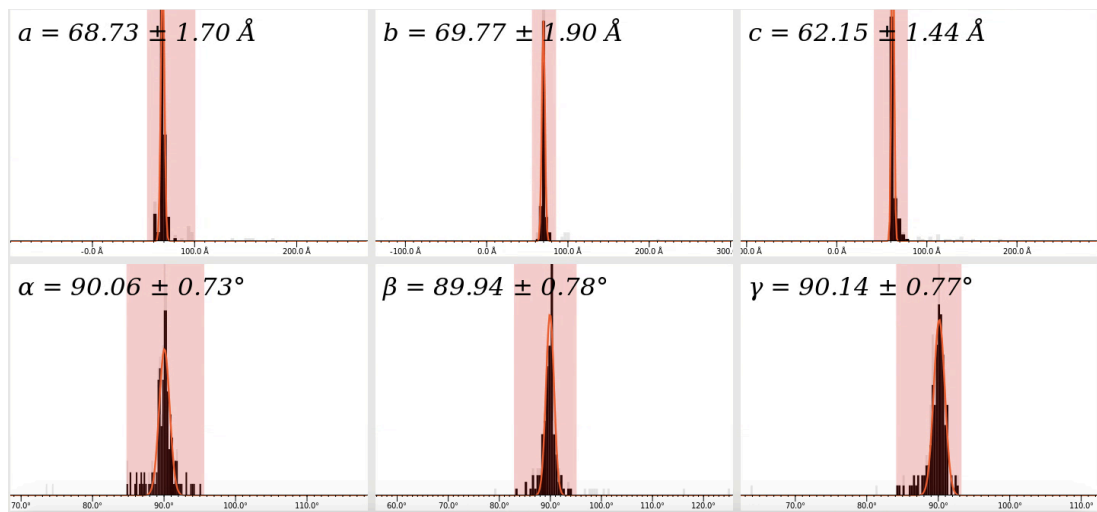


Figure 5.1 Histograms generated from *cell_explorer* for chip Fostis containing DHP crystals in complex with peracetic acid. The standard deviation and mean were calculated from the range defined by the highlighted area

The cell parameters were then updated into a new file called *DHP_refined.cell*. A second, full run of *indexamajig* was then carried out using the new refined cell parameters. The output of this run and the geometry file were used to check and optimize the detector geometry using *detector-shift*. This also plots the disagreement between the detector shift in x and y, between the location of the experimental and calculated spots in every indexed frame. To do this, this command was used

```
$ ./detector-shift DHP-PA-fostis.out new_geometry-  
predrefine.geom
```

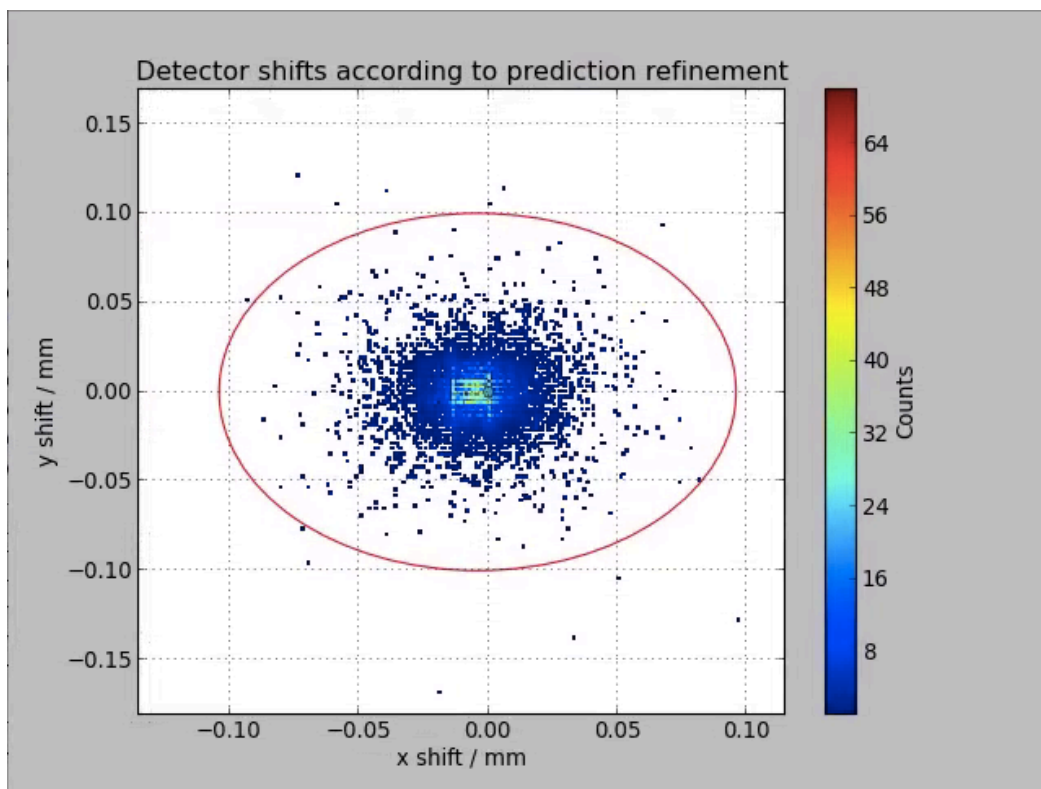


Figure 5.2 *Detector_shift* plot from CrystFEL for chip Fostis. The prediction-refinement algorithm refines the position of the beam on each frame, shown as a scatter plot

A new geometry file was made, called 'new_geometry-predefine-predefine' and a final run of *indexamajig* was done with the new geometry file. Cell explorer was

used to check that the histograms produced were much cleaner than the initial run. This is shown in Figure 5.3

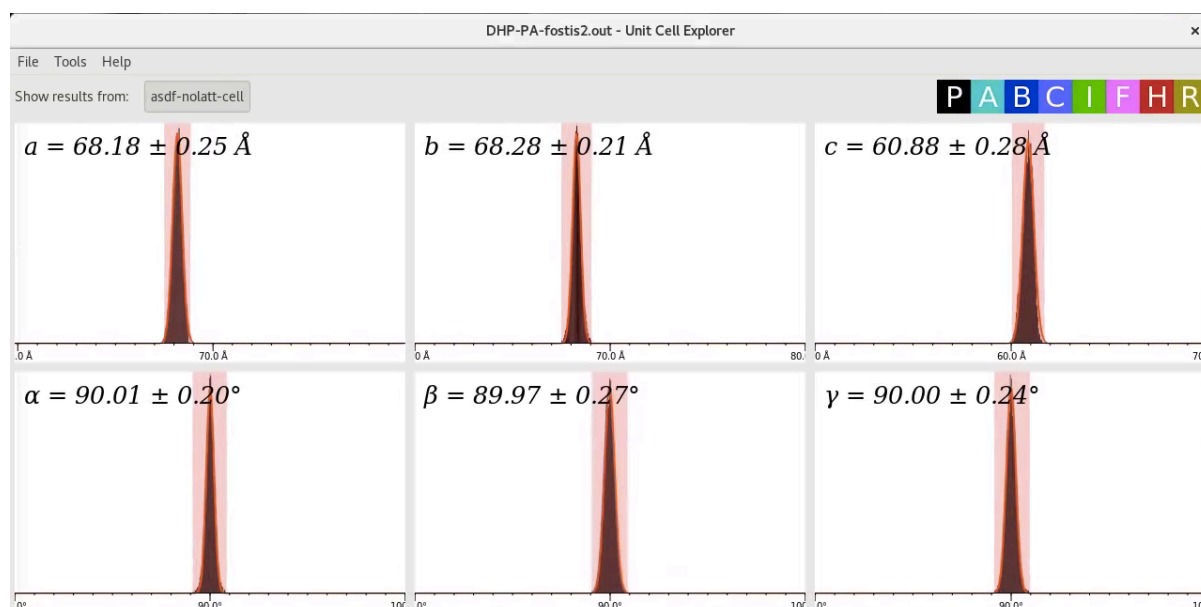


Figure 5.3: Histograms generated by Cell_explorer for chip ‘Foster’ after the final run of *indexamajig*

After, the out file was checked for detector saturation using the script ‘peakogram-stream’ the command line for this is

```
$ ./peakogram-stream -i DHP-PA-fostis2.out
```

Detectors at FEL facilities have a small dynamic range, and so there is a high possibility of ‘saturated reflections’. The graph produced from ‘peakogram-stream’ is shown in Figure 5.3. The vertical axis shows the highest pixel value in each reflection and the horizontal axis represents the resolution. Each point represents the density points, and there is a colour scale to show the density of points in areas that are very concentrated

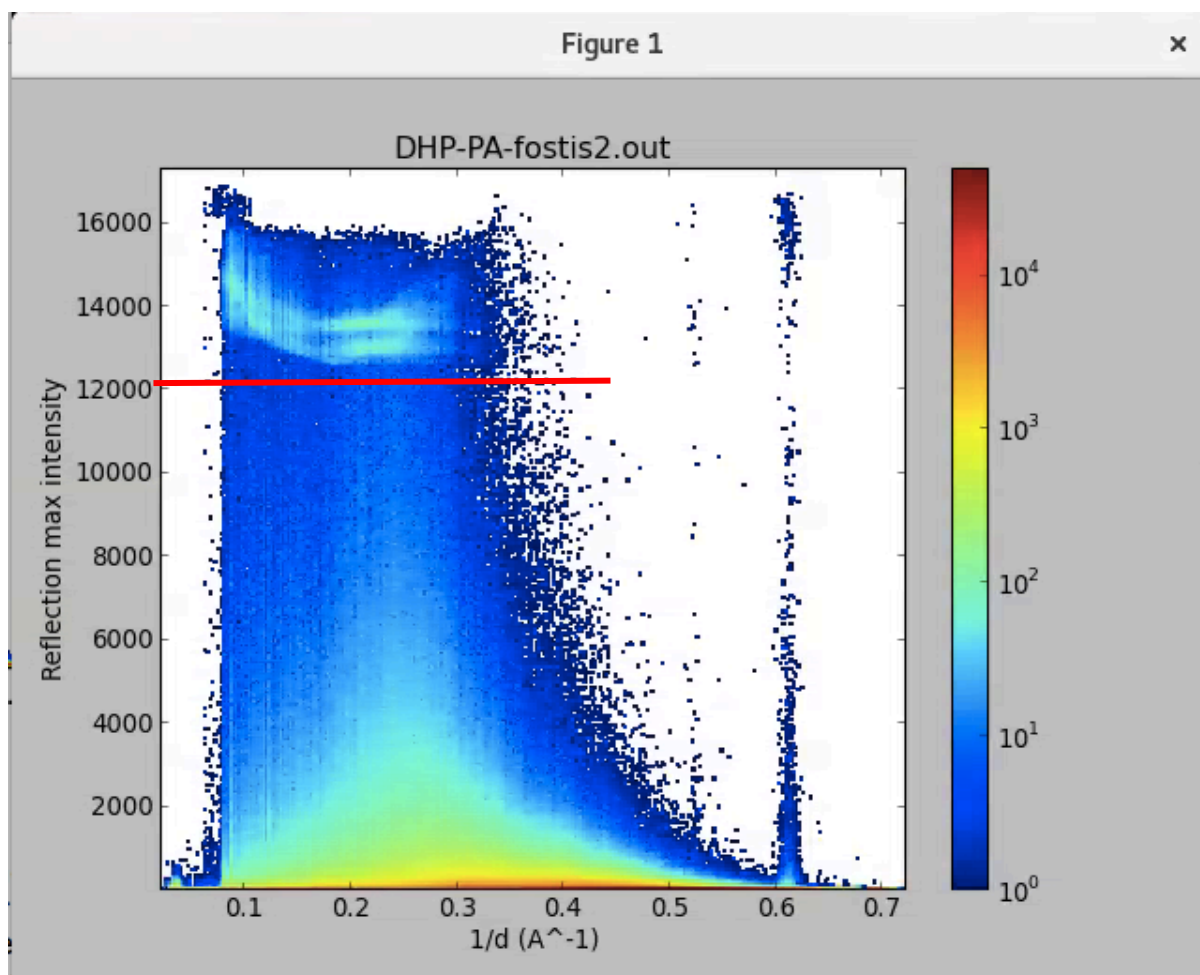


Figure 5.3 'peakogram-stream' graph. Each point represents one reflection, with the vertical axis showing the highest pixel value and the horizontal axis showing the resolution. The colour scale shows the density of points. The red line represents the intensity cut off

Due to the 'cloud' of density above the red line shown in Figure 5.3, values peaking over 12000 detector units are rejected when merging. The 'cloud' suggests reflections with a higher intensity than could be measured by the detector, and so they get clustered around the maximum value.

Next, the out files for all the different chips are joined together into one file ready to for scaling and merging. The command for this is:

```
$ cat DHP-PA-fostis2.out DHP-PA-ginger2.out DHP-PA-endive2.out >
DHP-PA-all.out
```

Merging was carried out using the simple, averaging-based 'Monte Carlo' method. To do this, *process_hkl* (White *et al.*, 2012) was used. This was given as a script

called *merge.sh* which was run with the output file DHP-PA-all.out, containing the outfiles from all *indexamajig* runs with the three different chips. This *merge.sh* file is shown in Figure 5.4 The point group is also needed in this script, which would be 222 as the space group for DHP is $P2_12_12_1$. However, with recommendation from CrystFEL developers, the point group 'mmm' was used to merge the Friedel pairs and therefore perform better merging.

Merging using *merge.sh* produces three output files, a .hkl file with all merged reflections, and the two files .hkl1 and hkl2, where the data was split. These split files are used to generate the useful figures of merit, CC and R_{split} .

To calculate these figures of merit, the script 'stat.sh' was used, written by Danny Axford. This requires an input of the symmetry information for DHP (DHP.cell) as well as the hkl files. The point group is specified as 'mmm' generating the correlation coefficient (CC) and R_{split} . The script for stat.sh is

```
#!/bin/bash

inp=$1
inp1=${1}1
inp2=${1}2
basename=DHP-PA-all.processhkl
fom="R1I R2 Rsplit CC CCstar"
pdb=DHP.cell
pg="mmm"
highres="1.85

module load CrystFEL

if ( ! -d "stat" ) then
    mkdir stat

fi

for mode in $fom
```

```

do
    compare_hkl $inpl $inp2 -y $pg -p $pdb --fom=$mode -
    highres=$highres --nshells=20 --shell-file='stat/${basename-
    $mode".dat 2>>stat/${basename}.log
done
check_hkl -p $pdb --nshells=20 --highres=$highres -y $pg -shell-
files="stat/${basename}-shells".dat $inp 2>>stat/${basename}.log

```

Each file generated by 'stat.sh' were 'dat' files. Shells.dat contains information about completeness and other data quality statistics in the different resolution shells. DHP data was completed using the resolution cut off 1.85 Å (Figure 5.6). The completeness did tail off from ~2.14 Å.

Center 1/nm	# refs	Possible	Compl	Meas	Red	SNR	Std dev	Mean	d(Å)	Min 1/nm	Max 1/nm
1.084	2051	2051	100.00	2876425	1402.5	15.97	1624.03	2270.01	9.23	0.176	1.992
2.251	1917	1917	100.00	1791702	934.6	15.25	1963.83	2356.52	4.44	1.992	2.509
2.691	1908	1908	100.00	1519597	796.4	13.45	1758.96	1633.56	3.72	2.509	2.872
3.017	1886	1886	100.00	1340644	710.8	11.63	1373.57	1012.73	3.31	2.872	3.161
3.283	1878	1878	100.00	1224862	652.2	10.04	1007.82	666.97	3.05	3.161	3.405
3.512	1871	1871	100.00	1121339	599.3	8.64	726.29	470.02	2.85	3.405	3.619
3.714	1849	1849	100.00	1027336	555.6	7.70	535.33	333.98	2.69	3.619	3.809
3.896	1869	1869	100.00	960437	513.9	6.78	379.49	224.38	2.57	3.809	3.983
4.063	1849	1849	100.00	878324	475.0	5.92	281.91	171.02	2.46	3.983	4.142
4.216	1853	1856	99.84	814994	439.8	5.31	208.80	121.83	2.37	4.142	4.290
4.360	1769	1858	95.21	750016	424.0	4.94	187.91	106.56	2.29	4.290	4.429
4.494	1658	1824	90.90	684720	413.0	4.56	124.77	77.74	2.23	4.429	4.559
4.621	1620	1846	87.76	645600	398.5	4.31	129.63	70.76	2.16	4.559	4.682
4.741	1555	1864	83.42	602872	387.7	3.94	101.76	55.39	2.11	4.682	4.800
4.855	1468	1835	80.00	548192	373.4	3.48	70.35	40.15	2.06	4.800	4.911
4.965	1406	1826	77.00	511052	363.5	3.22	59.77	34.42	2.01	4.911	5.018
5.069	1376	1841	74.74	488559	355.1	2.83	45.14	25.26	1.97	5.018	5.120
5.170	1332	1849	72.04	451490	339.0	2.12	37.06	20.47	1.93	5.120	5.219
5.266	1279	1843	69.40	425925	333.0	2.30	40.16	20.10	1.90	5.219	5.314
5.360	1213	1828	66.36	391681	322.9	2.06	33.87	16.38	1.87	5.314	5.405

Figure 5.6: shells.dat file for DHP in complex with peracetic acid with a cut off of 1.85 Å

To further look into the quality of this data and to figure out the resolution cut off, the stats CC and R_{split} must be looked at. The CC statistics are shown in Figure 5.7

1/d centre	CC	nref	d / Å	Min 1/nm	Max 1/nm
1.084	0.9830890	2051	9.22	0.177	1.992
2.251	0.9843278	1917	4.44	1.992	2.509
2.691	0.9870889	1908	3.72	2.509	2.872
3.017	0.9867705	1886	3.31	2.872	3.161
3.283	0.9854015	1878	3.05	3.161	3.405
3.512	0.9830649	1871	2.85	3.405	3.619
3.714	0.9831473	1849	2.69	3.619	3.809
3.896	0.9677920	1870	2.57	3.809	3.983
4.063	0.9743446	1848	2.46	3.983	4.142
4.216	0.9762509	1839	2.37	4.142	4.290
4.360	0.9786313	1755	2.29	4.290	4.429
4.494	0.9585517	1640	2.23	4.429	4.559
4.621	0.9682513	1604	2.16	4.559	4.682
4.741	0.9571053	1540	2.11	4.682	4.800
4.855	0.9403951	1458	2.06	4.800	4.911
4.965	0.9359669	1392	2.01	4.911	5.018
5.069	0.7603989	1365	1.97	5.018	5.120
5.170	0.8436649	1315	1.93	5.120	5.219
5.266	0.8504675	1266	1.90	5.219	5.314
5.360	0.5681491	1204	1.87	5.314	5.405

Figure 5.7: shells.dat file for DHP in complex with peracetic acid with a cut off of 1.85 Å generated by the Stat.sh script. Correlation coefficient data is shown.

1/d centre	Rsplit/%	nref	d / Å	Min 1/nm	Max 1/nm	, 20%,
1.084	6.32	2051	9.22	0.177	1.992	
2.251	6.84	1917	4.44	1.992	2.509	,
2.691	7.16	1908	3.72	2.509	2.872	
3.017	8.24	1886	3.31	2.872	3.161	
3.283	8.75	1878	3.05	3.161	3.405	
3.512	9.58	1871	2.85	3.405	3.619	
3.714	10.44	1849	2.69	3.619	3.809	
3.896	12.05	1870	2.57	3.809	3.983	
4.063	12.54	1848	2.46	3.983	4.142	
4.216	13.51	1839	2.37	4.142	4.290	
4.360	13.83	1755	2.29	4.290	4.429	
4.494	16.60	1640	2.23	4.429	4.559	
4.621	17.22	1604	2.16	4.559	4.682	
4.741	19.10	1540	2.11	4.682	4.800	
4.855	24.51	1458	2.06	4.800	4.911	
4.965	26.31	1392	2.01	4.911	5.018	
5.069	32.25	1365	1.97	5.018	5.120	
5.170	37.52	1315	1.93	5.120	5.219	
5.266	40.91	1266	1.90	5.219	5.314	
5.360	49.53	1204	1.87	5.314	5.405	

Figure 5.8: shells.dat file for DHP in complex with peracetic acid with a cut off of 1.85 Å generated by the Stat.sh script. R_{split} data is shown.

To see if the data improved, *partialator*, part of the CrystFEL suite was run.

Partialator performs advanced merging methods, implements scaling and post refinement using generalised target function, a choice of partiality models, cross-validation residuals and partiality plots while 'process_hkl' is the simplest way to merge intensities, by calculating the average intensity across the entire data set for each symmetrically unique reflection (White 2014). *Partialator* was run using the command:

```
$ partialator -i DHP-PA-all.out -o DHP-PA-all.partialator.hkl -y  
mmm --iterations=3 --model=unity -max-adu=7000 -j 20
```

The figures of merit were then calculated in the same way as previously stated for *process_hkl*. A table with a comparison of the figures of merit of *process_hkl* vs *partialator* is shown in Table 5.5. With no significant improvement in any of the figures of merit, the merging from *process_hkl* was used.

Table 5.6: Figures of merit for merging processes *process_hkl* and *partialator*

Figure of Merit	<i>Process_hkl</i>	<i>Partialator</i>
R _{split} (%)	8.12	9.17
CC	0.991	0.996
I/ σ (I)	11.48	12.9

An mtz file was generated using the script 'create-mtz' which runs the CCP4 program *f2mtz*. The hkl file from merging was used as the input file. The script is shown:

```
#!/bin/sh
```

```
OUTFILE= echo $1 | sed -e 's/\.hkl$/\.mtz/'
```

```

echo = Input: $1"
echo =Output: $OUTFILE"
if [ -e $OUTFILE ]; then
    echo = The output file already exists
    echo = -$OUTFILE
    echo = To confirm that you want to continue, which
will DESTROY the"
    echo = current contents of this file, type y and press
enter."
    read conf
    if [ $conf !=y] then
        echo = "Not confirmed"
        exit 1
    else
        echo = "Proceeding"
    fi
fi

ex -c /End of reflections/
..$d
w create-mtz.temp.hkl
q!' $1

Echo "Running `f2mtz`
f2mtz HKLIN create-mtz.temp.hkl HKLOUT $OUTFILE > out.html <<
EOF
TITLE Reflections from CrystFEL
NAME PROJECT SACLA CRYSTAL DHP-PA DATASET EMILY allchips
CELL 68.73 69.77 62.15 90 90 90
SYMM P212121
SKIP 3
LABOUT H K L IMEAN SIGIMEAN
CTYPE H H H J      Q
FORMAT `(3(F4.0.1X),F10.2.10X.F10.2)'
EQF

if [ $? -ne 0]; then echo "Failed,"; exit; fi
rm -f create-mtz.temp.hkl

```

echo "done"

The resulting mtz file contained the space group and experimental parameters specified previously for DHP and peracetic acid. The mtz was then used the same way as other datasets, starting with molecular replacement using *Phaser*, using a room-temperature SFX model of Ferric DHP from a previous SACLA trip. Data quality statistics for DHP Peracetic acid is shown in Table 5.6

Table 5.7: Refinement statistics and data collection for SFX data measured at SACLA of DHP-Peracetic Acid. Values in brackets represent the stats from the highest shell of resolution

DHP-Peracetic Acid	
Space group	P2 ₁ 2 ₁ 2 ₁
Unit cell (Å)	a=68.73 b=69.77 c=62.15 $\alpha=\beta=\gamma=90^\circ$
Resolution (Å)	30.55-1.85
R _{split} (%)	8.12 (59.99)
Unique reflections	22935 (2429)
I/ σ (I)	11.48 (8.94)
CC	0.991 (0.56)
Completeness (%)	99.97 (100)
R _{work}	0.23
R _{free}	0.25
Reflections for R-free	834 (78)
Rmsd bond lengths (Å)	0.019
Rmsd bond angles (°)	1.39
Ramachandran	
Favoured (%)	98.15

5.4 SFX Structure of DHP with Peracetic Acid (formation of Complex I)

The data for Peracetic acid soak with DHP-B was measured at SACLA, Japan in 2018 and the data processed by me. This structure was measured using a 10 femtosecond XFEL pulse to collect the radiation-damage free structure. The data was processed as previously stated, using CrystFEL.

The structure, measured at room temperature, crystallized as a dimer in the asymmetric unit (Figure 5.9) with each monomer showing the globin fold as expected. The globin fold is formed by seven α -helices per monomer, bundled around the haem co-factor.

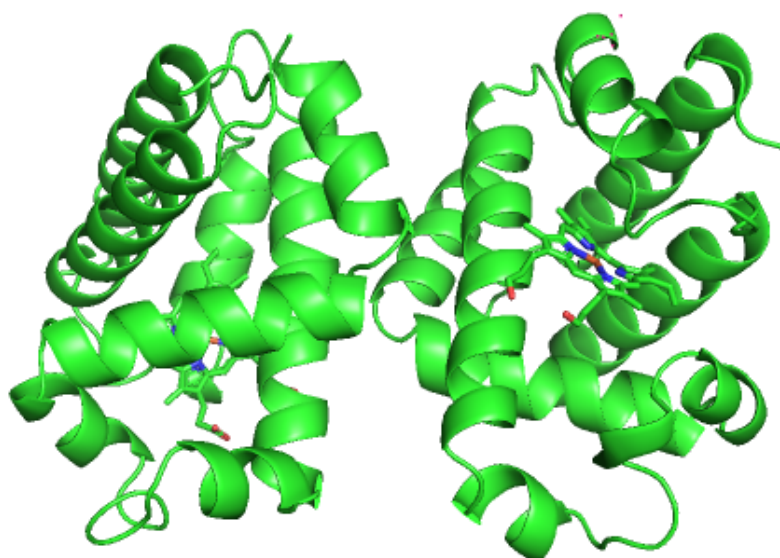


Figure 5.9 Overall structure of the DHP dimer made using the SFX structure measured at SACLA in 2018. Haem groups are shown as sticks, with monomer A shown on the left, and B on the right.

The complex was solved using molecular replacement using *Phaser*, and after one round of refinement using *REFMAC5*, a large positive electron-density in the $F_o - F_c$ difference map was seen just above the haem plane. Due to this peracetic acid soak's aim of oxidising the haem to compound one, an water molecule was modelled in. After refinement, the water was only supported by electron density at haem B, and so was removed at haem A.

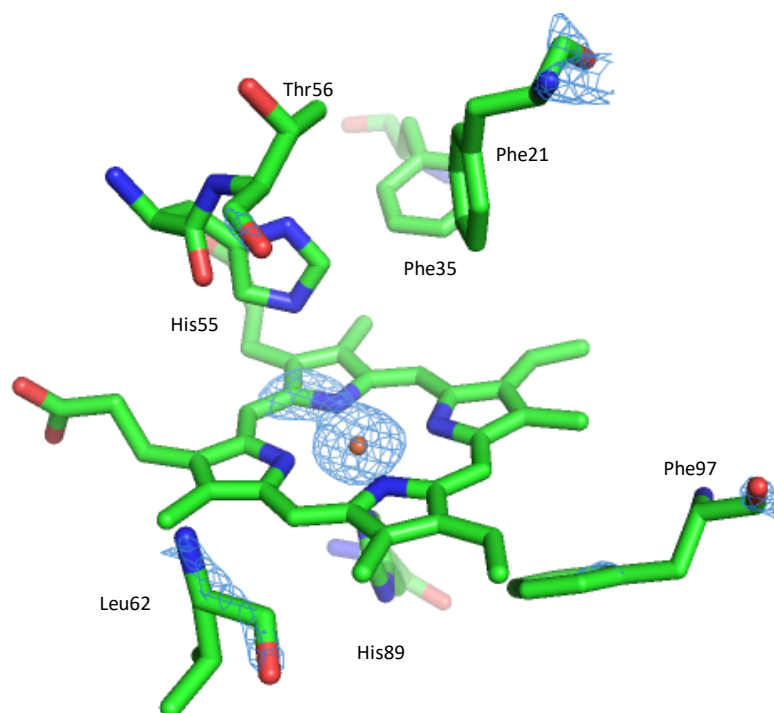


Figure 5.11 $2F_o - F_c$ map of Haem A of DHP from the first round of refinement contoured to 4σ to show the two positions of the haem iron in the hemichrome species

Haem A appears to have a hemichrome species, a double conformation of the haem, and has previously been observed in other haem proteins (Vergara *et al.*, 2009) shown in Figure 5.11 with the electron density map contoured to 4σ to show the two positions of the iron more clearly. Due to the data being an XFEL structure, radiation damage being the reasoning behind seeing this hemichrome species is ruled out. This has been observed in other similar structures, so currently it is unsure whether this has any biological relevance or just a crystal artefact.

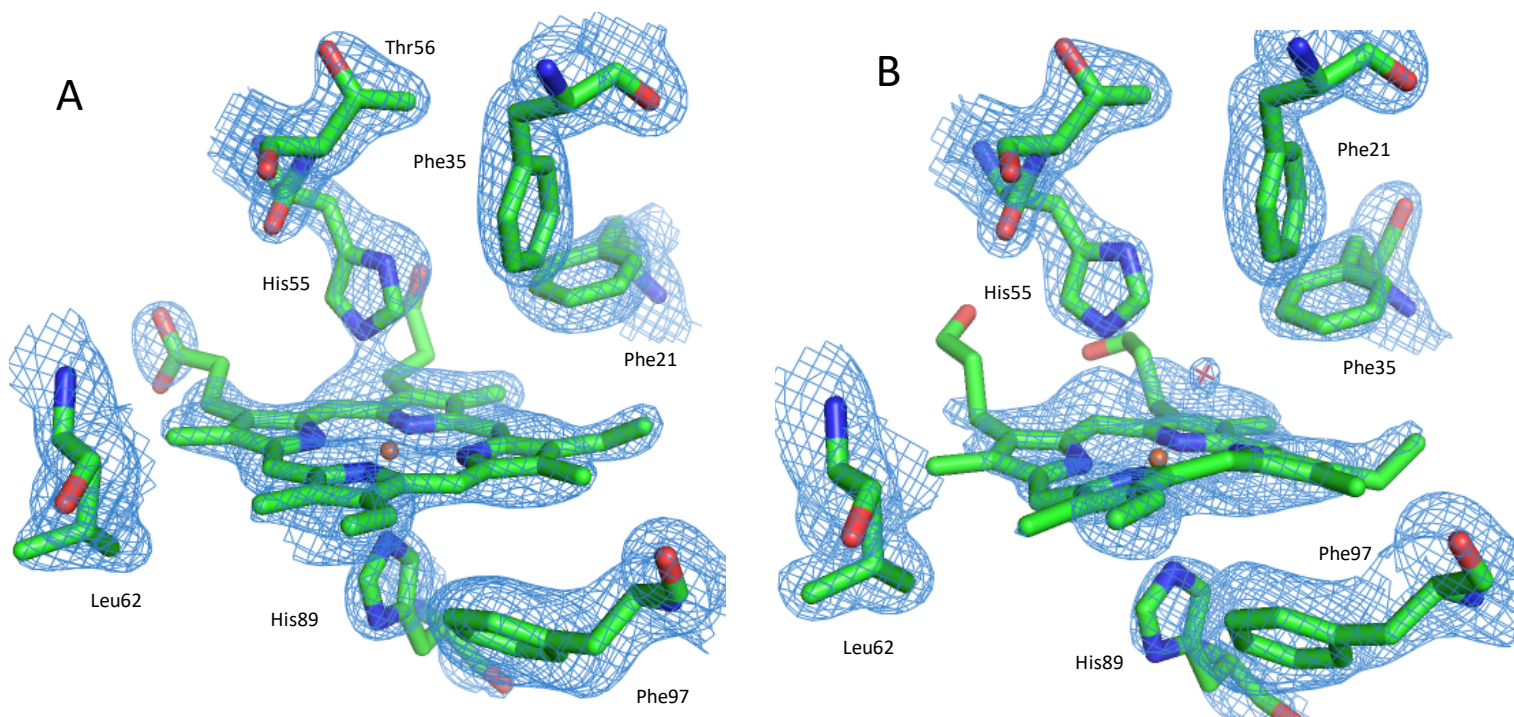


Figure 5.12 Haem A (A) and Haem B (B) electron density contoured to 1σ . An water molecule was modelled into the electron density above the haem plane in haem B.

The B-factor and occupancy data for haem B is shown in Table 5.8. B-factors reflect the fluctuations of atoms about their average position, providing details on protein dynamics (Yuan, Bailey and Teasdale, 2005)

Table 5.8: B-factor and occupancy details for haem B of DHP-B peracetic acid

	B- factor (\AA^2)	Occupancy (%)
Water	45.73	70
Fe	56.61	100

5.4.1 Comparison of Ferric DHP-B and DHP-B and Peracetic Acid

A comparison of the structure of DHP-B and peracetic acid was compared to the ferric state DHP-B structure collected from a previous SACLA trip (Moreno-Chicano

et al 2019) . The structures are extremely similar, with very little differences. At haem A, the biggest difference that can be observed is of the distal histidine. In the PA soak, the distal histidine is 4.23Å away from the iron of the haem group, whereas in the Ferric structure it is 4.04 Å at haemA. Although closer to the haem iron, the histidine in the PA soak is much closer to the haem plane itself as seen in Figure 5.13, while in the Ferric form His55 is directly above the iron. At haem B, the distal histidine is slightly closer to the haem iron than in the ferric In haem B, being 0.4 Å closer. Also at haem B, the haem plane is in a ‘domed’ position.

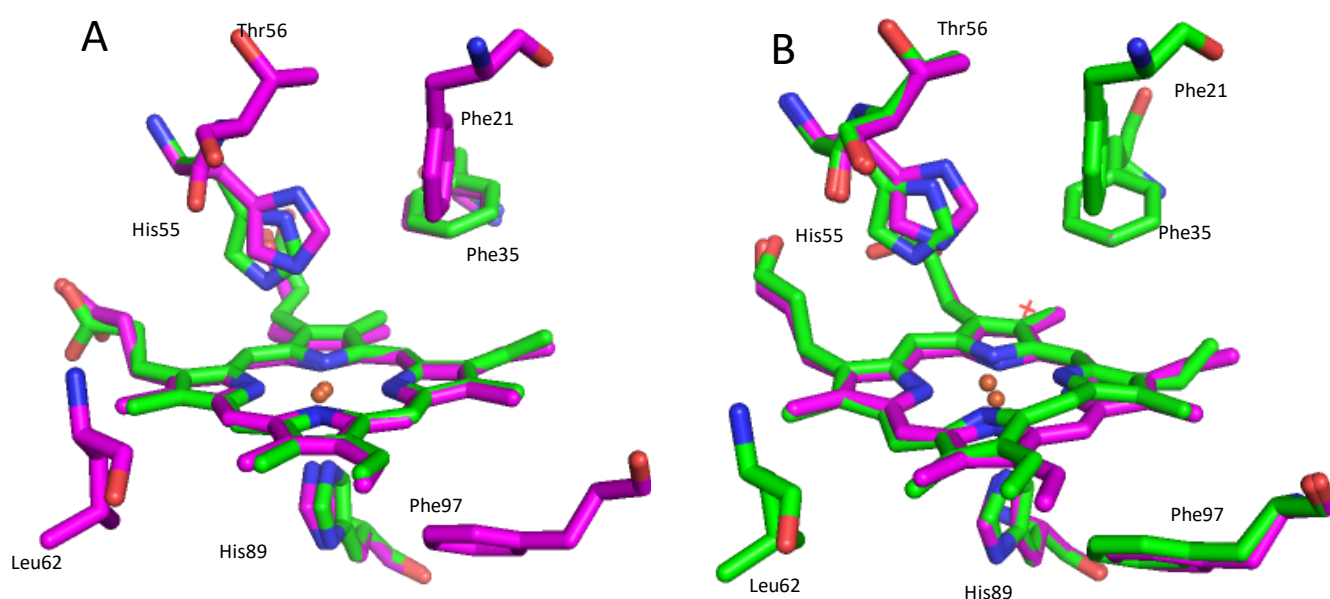


Figure 5.13 Superstition of SFX structures Haem A (A) and Haem B (B) of DHP-B with Peracetic Acid (Green) and Ferric DHP-B (Pink).

Table 5.9: Key residue distances for DHP-B SFX structures Ferric and peracetic acid soak

Structure	His55 – Fe Distance (Å)	Fe – His89 Distance (Å)	Phe35 Angle (°)
Ferric – haem A	4.04	2.51	80
Peracetic acid -Haem A	4.23	2.45	114
Ferric – haem B	4.86	2.48	111
Peracetic acid – Haem B	4.44	2.88	113

There is also a Phe35 rotation, as shown by the angles in Table 5.9. Ferric DHP has Phe34 at 80 degrees at haem A, while the PA soak is 114 degrees, a 34 degree turn.

5.5 SFX Structure of Oxyferrous DHP

Oxyferrous DHP-B was generated by incubating ferric DHP with ascorbic acid, and microcrystals grown as previously stated in Chapter 3. The oxyferrous data was measured at SACLA, Japan in 2018 by Dr Tadeo Moreno-Chicano, and the data processed by me. The oxyferrous data was collected using a 10 femtosecond XFEL pulse, this ensured a radiation-damage free structure, the same way the peracetic acid structure was obtained. The data was processed in the same way as previously stated, using CrystFEL following the same workflow as DHP with peracetic acid. The chips were indexed with *indexamajig* using the same parameters as before. The diffraction data was joined into a single file, containing 55,431 reflections, with the cell parameter distributions shown in Figure 5.10 calculated using *cell_explorer*

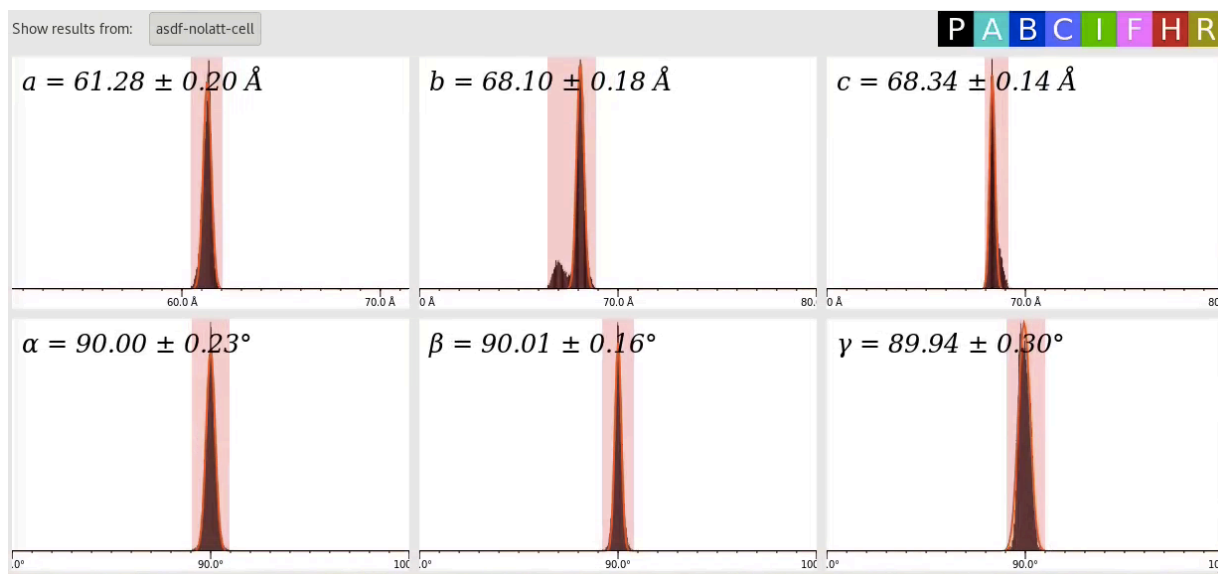


Table 5.12: Histogram plots generated by *cell_explorer* for DHP-oxyferrous data.

The merging processes was performed by *partialator* and the data quality statistics were generated using the script *stat.sh*. Resolution shells for this complex are shown in Figure 5.11, the CC statistics in Figure 5.12 and R_{split} in Figure 5.13.

Center	1/nm	# refs	Possible	Compl	Meas	Red	SNR	Std dev	Mean	d(A)	Min 1/nm	Max 1/nm
1.105	1389	1389	100.00	1045149	752.4	9.71	950.63	925.55	9.05	0.218	1.992	
2.251	1292	1292	100.00	661887	512.3	10.86	984.76	1233.40	4.44	1.992	2.509	
2.691	1299	1299	100.00	574559	442.3	10.82	712.46	917.41	3.72	2.509	2.872	
3.017	1263	1263	100.00	501167	396.8	10.46	413.51	533.64	3.31	2.872	3.161	
3.283	1256	1256	100.00	432280	344.2	9.95	295.10	363.28	3.05	3.161	3.405	
3.512	1266	1266	100.00	379515	299.8	8.99	180.14	238.79	2.85	3.405	3.619	
3.714	1246	1246	100.00	375868	301.7	8.58	125.95	163.27	2.69	3.619	3.810	
3.896	1258	1258	100.00	369223	293.5	8.36	102.03	134.94	2.57	3.810	3.983	
4.063	1229	1229	100.00	354550	288.5	7.99	87.20	116.14	2.46	3.983	4.142	
4.216	1266	1266	100.00	351951	278.0	7.19	65.48	83.71	2.37	4.142	4.290	
4.360	1235	1235	100.00	336477	272.5	6.75	55.78	71.28	2.29	4.290	4.429	
4.494	1215	1215	100.00	325130	267.6	6.41	43.57	57.97	2.23	4.429	4.559	
4.621	1245	1245	100.00	321154	258.0	5.70	36.17	45.89	2.16	4.559	4.682	
4.741	1241	1241	100.00	313589	252.7	4.89	33.78	34.96	2.11	4.682	4.800	
4.855	1239	1239	100.00	305875	246.9	4.46	24.65	28.97	2.06	4.800	4.911	
4.965	1242	1242	100.00	300852	242.2	3.75	19.38	20.99	2.01	4.911	5.018	
5.069	1226	1226	100.00	290924	237.3	3.29	15.28	16.53	1.97	5.018	5.120	
5.170	1240	1240	100.00	286671	231.2	2.70	11.98	12.70	1.93	5.120	5.219	
5.266	1239	1239	100.00	281484	227.2	2.37	10.70	10.37	1.90	5.219	5.314	
5.360	1213	1213	100.00	271268	223.6	2.11	8.10	8.39	1.87	5.314	5.405	

Figure 5.13 Resolution shells for the merged data for oxyferrous DHP

1/d centre	CC	nref	d / A	Min 1/nm	Max 1/nm
1.106	0.9656371	1389	9.04	0.219	1.992
2.251	0.9572773	1292	4.44	1.992	2.509
2.691	0.9375728	1299	3.72	2.509	2.872
3.017	0.9434513	1263	3.31	2.872	3.161
3.283	0.9469602	1256	3.05	3.161	3.405
3.512	0.9321255	1266	2.85	3.405	3.619
3.714	0.9356972	1246	2.69	3.619	3.810
3.896	0.9354480	1258	2.57	3.810	3.983
4.063	0.9264751	1229	2.46	3.983	4.142
4.216	0.9215566	1266	2.37	4.142	4.290
4.360	0.9165281	1235	2.29	4.290	4.429
4.494	0.8958585	1215	2.23	4.429	4.559
4.621	0.8885901	1245	2.16	4.559	4.682
4.741	0.8979949	1241	2.11	4.682	4.800
4.855	0.8246448	1239	2.06	4.800	4.911
4.965	0.8121769	1242	2.01	4.911	5.018
5.069	0.7620005	1226	1.97	5.018	5.120
5.170	0.6429644	1240	1.93	5.120	5.219
5.266	0.5642558	1239	1.90	5.219	5.314
5.360	0.4992503	1213	1.87	5.314	5.405

Figure 5.14 Values for CC In the different resolution shells for the merged data for oxyferrous DHP

1/d centre	R _{split} /%	nref	d / Å	Min 1/nm	Max 1/nm
1.106	10.58	1389	9.04	0.219	1.992
2.251	10.67	1292	4.44	1.992	2.509
2.691	11.09	1299	3.72	2.509	2.872
3.017	11.67	1263	3.31	2.872	3.161
3.283	11.38	1256	3.05	3.161	3.405
3.512	12.74	1266	2.85	3.405	3.619
3.714	13.05	1246	2.69	3.619	3.810
3.896	13.32	1258	2.57	3.810	3.983
4.063	13.74	1229	2.46	3.983	4.142
4.216	14.85	1266	2.37	4.142	4.290
4.360	15.64	1235	2.29	4.290	4.429
4.494	17.23	1215	2.23	4.429	4.559
4.621	18.64	1245	2.16	4.559	4.682
4.741	21.63	1241	2.11	4.682	4.800
4.855	25.62	1239	2.06	4.800	4.911
4.965	30.96	1242	2.01	4.911	5.018
5.069	36.23	1226	1.97	5.018	5.120
5.170	45.44	1240	1.93	5.120	5.219
5.266	53.49	1239	1.90	5.219	5.314
5.360	59.99	1213	1.87	5.314	5.405

Figure 5.15 Values for R_{split} in the different resolution shells for the merged data for oxyferrous DHP

The data quality/merging statistics are shown in Table 5.5. The structure was initially solved using *Phenix* however the CCP4i2 suite gave a clearer definition of the oxygen bound to the haem. This is due to the fact that the crystals were twinned (twin law - h,l,k) , and *REFMAC5* (Murshudov *et al.*, 2011) dealt with twinned data better than in *Phenix*, resulting in clearer maps. The structure was solved by doing an initial molecular replacement using *PHASER* (McCoy *et al.*, 2007) the model built using *BUCCANEER* (Cowtan, 2006) and refined using *REFMAC5* as well as manual refinement in *Coot*. Due to the twinned data, the recommended *LORESTR* (Kovalevskiy, Nicholls and Murshudov, 2016) was also used for refinement.

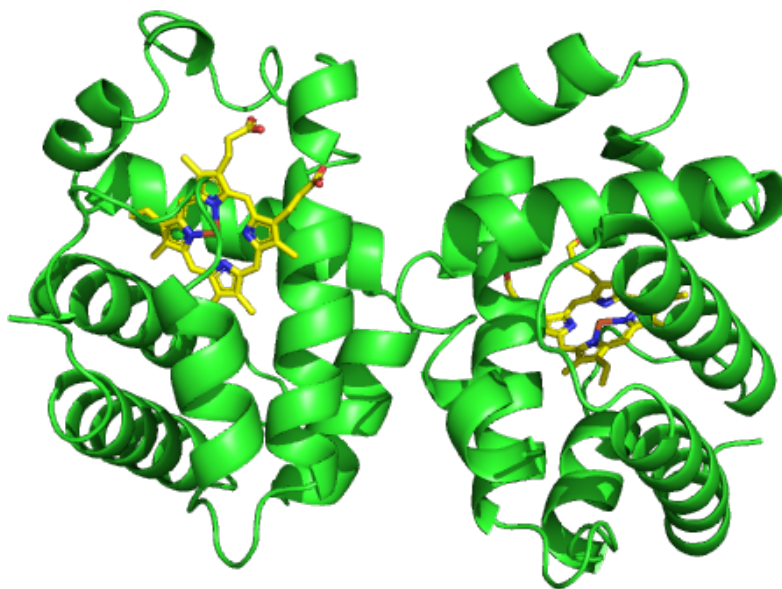


Figure 5.16 Overall structure of the oxyferrous DHP dimer made using the SFX structure measured at SACLA. Haem groups are represented in yellow.

The SFX structure for oxyferrous DHP shows that it crystallised as expected, as a dimer in the asymmetric unit, with each monomer adopting the classic globin fold. The haem groups can be distinctly seen. After the first round of refinement using REFMAC5, an area of positive electron density in the $F_o - F_c$ map was observed above both haem A and B in the distal pocket (Figure 5.17), where the oxygen molecules were modelled into. After further refinement, the oxygen at haem A was not supported by electron density, and so was removed from the model.

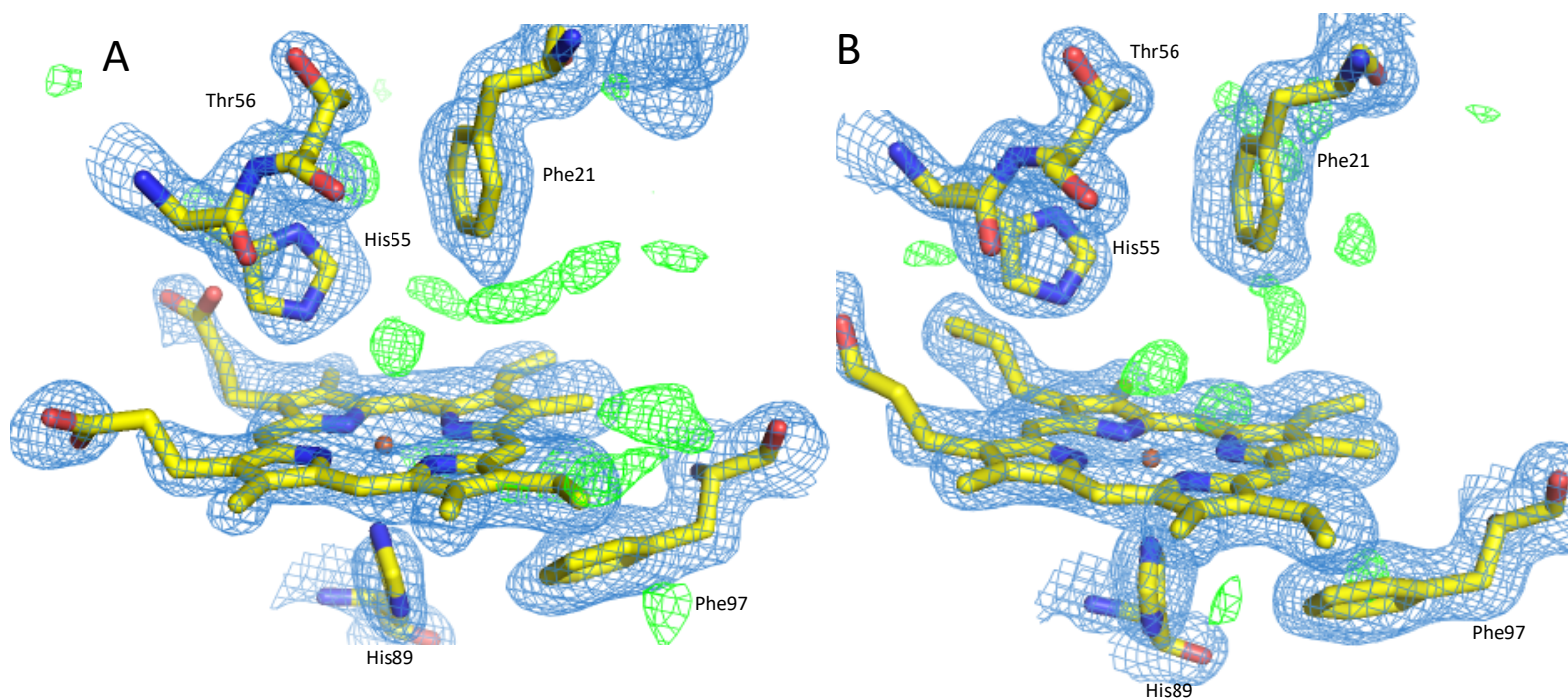


Figure 5.17 Haem A (A) and Haem B (B) from the first cycle of refinement. The difference F_o-F_c contoured to 3σ and density contoured to 1σ . Note the areas of positive density feature above the iron in the haem group

Other positive difference electron density clouds in the distal pocket are removed after multiple rounds of refinement, as this is noise in the data. Where the oxygen is modelled in, this area of electron density is apparent in all rounds of refinement.

In haem A, there is the presence of a hemichrome species. The His55 is also 1.34 \AA closer to the haem group in haem A. To more clearly show the hemichrome species, Figure 5.18 shows HaemA with $2F_o-F_c$ map contoured to 4σ , where the two positions of the iron can be clearly seen.

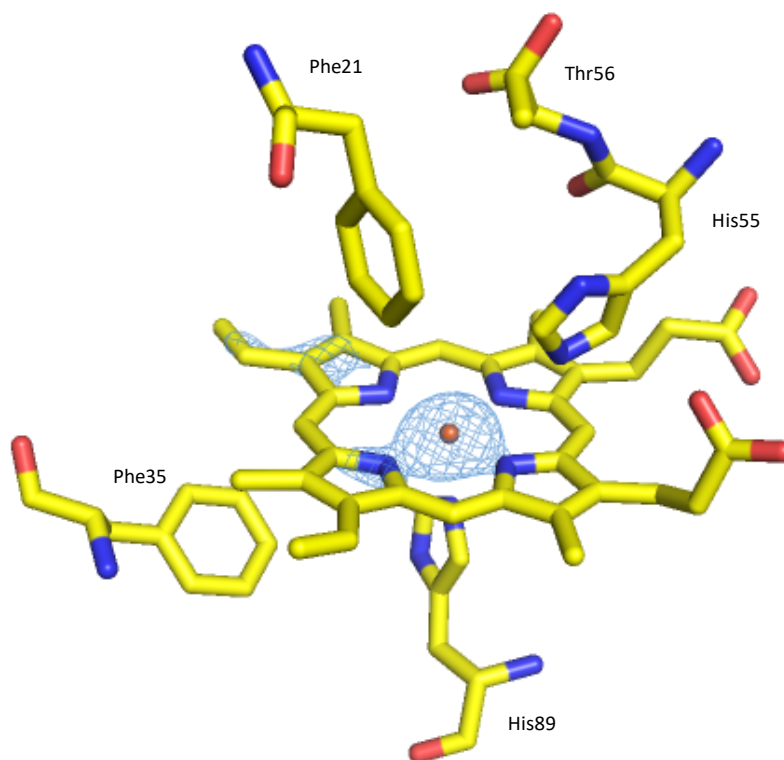


Figure 5.18 Haem A (A) from the first cycle of refinement in CCP4i2. F_o-F_c map contoured to 4σ to show the two positions of the haem iron, demonstrating a hemichrome species.

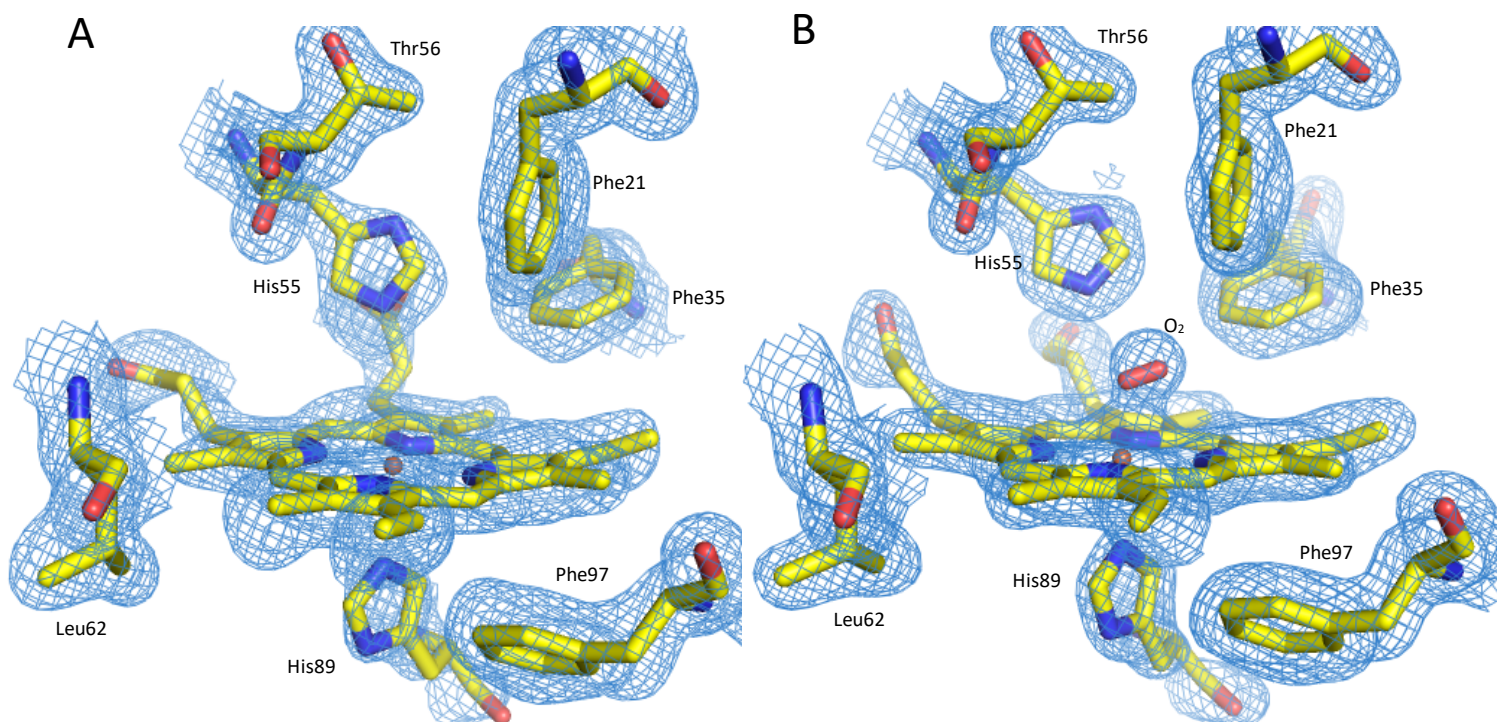


Figure 5.19 (A) Haem A of DHP-B, the O_2 was removed as the electron-density map did not support it. (B) Haem B, with the O_2 modelled in after multiple rounds of refinement. Map on both contoured to 1σ

The bent bonding geometry of the iron bonding to the O₂ molecule results in a hydrogen bond of length 3.14Å between the second O atom of molecular oxygen and the Ne2 of the distal histidine residue His55.

B-factor and occupancy information is shown in Table 5.9.

Table 5.10: B-factor and occupancy of the iron in the porphyrin and O₂ molecule in the active site of B

	B-factor (Å ²)	Occupancy (%)
Fe	34.51	100
O ₂	45.93	50

Table 5.11: Refinement statistics and data collection for SFX data measured at SACLA of Oxyferrous DHP. Values in brackets represent the stats from the highest shell of

	Oxyferrous
Space group	P2 ₁ 2 ₁ 2 ₁
Unit cell	a = 61.31 Å b= 68.1Å c=68.33 Å α=β=γ=90°
Resolution	11.34 – 2.0 Å
R _{split} (%)	12.03 (59.99)
Unique reflections	25099 (2653)
I/σ (I)	18.71 (13.12)
CC	0.975 (0.499)
Completeness	100% (100%)
R _{work}	0.20
R _{free}	0.22
Reflections for R-free	997 (144)
Rmsd bond lengths (Å)	0.024
Rmsd bond angles (°)	2.07
Ramachandran	
Favoured (%)	97.78
Twin law	-h,l,k

5.5.1 Comparison of DHP Oxyferrous – Room Temperature (SFX) and Cryogenic

As previously stated, crystallography is almost always performed at cryogenic temperatures (100K) to reduce radiation damage. Collecting at cryogenic temperatures however can mask alternative conformations that can be seen at room temperatures. Comparing the cryogenic structure (Carey *et al.*, 2018) and SFX

data collected in 2017 at SACLA (Japan). Haem B for both cryogenic and SFX data are shown in Figure 5.20

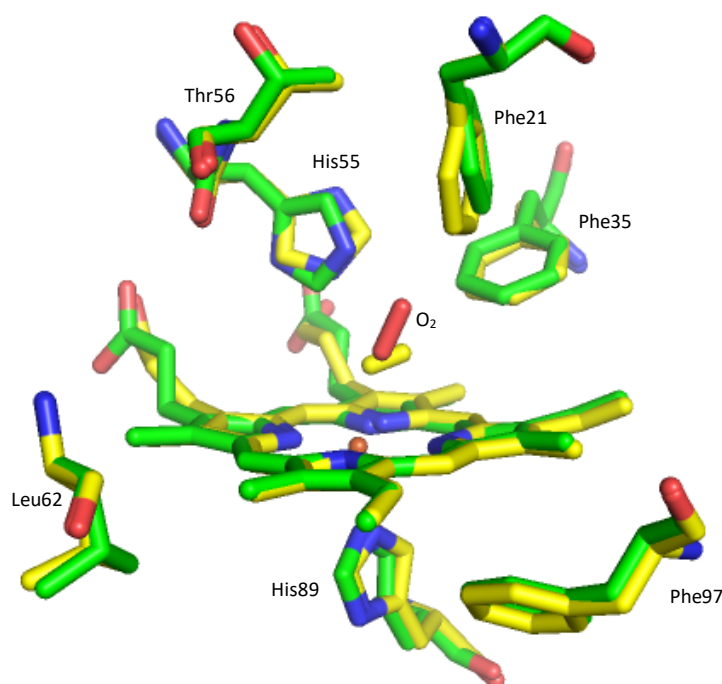


Figure 5.20 Superposed haem B from DHP-B. SFX data is shown in yellow and cryogenic shown in green

Comparisons of ligand to key residues are shown in Table 5.5. The cryogenic and SFX structures are quite similar from the comparisons of the structure shown in Figure 1.8, however looking at the key residue distances, the oxygen in the SFX structure is closer to the haem plane by 0.5 Å. The biggest difference is the distance between oxygen and the iron of the haem, due mainly to the oxygen being modelled in at a different angle. There is also a large difference between the angle at which the oxygen is bound. In the cryogenic, the oxygen is at a 169° angle to the iron, whereas the SFX data has the oxygen at 113°, almost parallel to the haem. Another slight difference is the distance between Phe21 and the oxygen. In the cryogenic, the Phe21 is closer to the oxygen at 3.35 Å, while in the SFX data, Phe21 is 3.88 Å. The structures are extremely similar, with the oxygen angle being the main difference.

Table 5.5: Key residue distances for oxyferrous DHP-B complexes for cryogenic and SFX structures

Complex	His55 - Ligand Distance (Å)	Fe - Ligand Distance (Å)	His89 - Ligand Distance (Å)
Cryogenic Oxyferrous	3.41	2.55	4.78
SFX Oxyferrous	3.24	2.05	4.32

5.6. Discussion

5.6.1. DHP Serial Crystallography

DHP has been investigated using serial crystallography at x-ray sources, such as XFEL and synchrotron, and has been proven to be an effective technique. SFX and SSX are efficient in many ways, including sample and time. In theory, with each chip containing 10,000 microcrystals running for 15 minutes, a complete data set could be obtained in around an hour.

This thesis shows the structure of two dehaloperoxidase-B structures, measured using SFX. Previously, three more DHP-B room temperature structures have been solved. Using room-temperature structures is advantageous, as the protein dynamics aren't suppressed so can show a more relevant model that is more biologically accurate. This is due to there being no cryoprotectants and having conditions that are closer to the *in vivo* conditions of DHP.

In dehaloperoxidase B peracetic acid soaked SFX structure, the overall structure is as expected, crystallising as a dimer in the asymmetric unit. There is also the globin fold formed by seven α -helices per monomer.

At haem B there is a clear positive electron density in the difference ($F_o - F_c$) map, showing full occupancy of a distal oxygen molecule and so could be showing the oxidation of the haem to compound one.

At haem A of DHP-B and peracetic acid, there is evidence of a hemichrome species. After the first round of refinement, the electron density viewed at 4σ showed the two possible positions of the haem iron. Further evidence of this is the closeness of the distal histidine to the haem plane, which is also a sign of a hemichrome species as the flexibility of His55 is what gives DHP the ability to form a hemichrome species (Nicoletti *et al.*, 2011)

5.6.2 Comparison between Ferric DHP and Peracetic Acid soaked DHP

DHP-B microcrystals were soaked in peracetic acid to oxidise the DHP to compound I. As previously shown in Chapter One, the suggested catalytic cycle of dehaloperoxidase shows a single oxygen atom bound to the haem iron as well as interacting with the NH of the distal histidine. An overall superposed structure of ferric DHP and DHP with a peracetic acid soak can be seen at Figure 5.21.

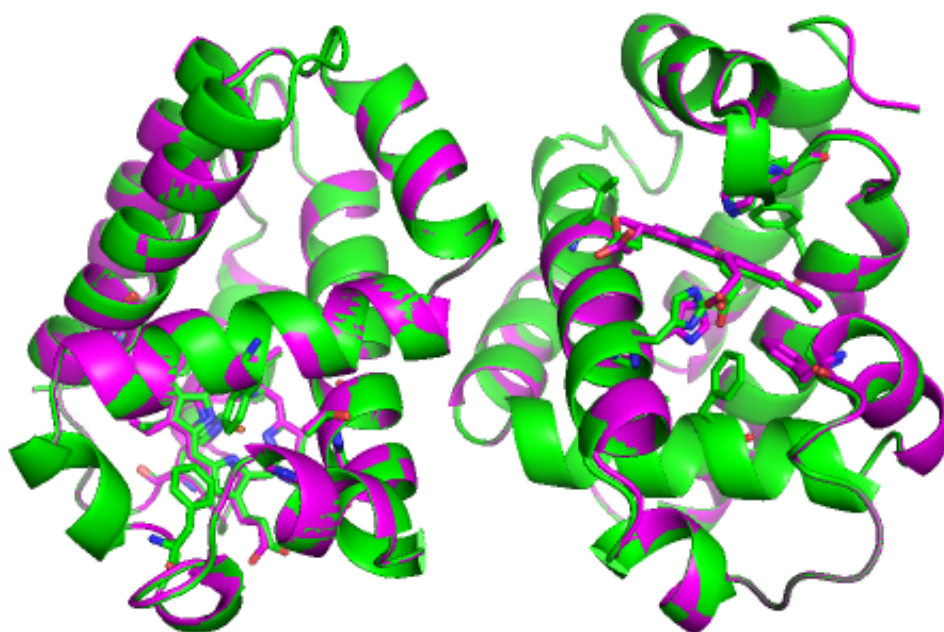


Figure 5.21. Superposition of DHP-B with peracetic acid soak (green) and Ferric structure (pink). Monomer A is shown on the left and monomer B on the right.

Apart from some differences at the haem group, they adopt almost identical conformations. The Rmsd (root-mean-square distance) between the two structures are shown in Table 5.6. Rmsd is the measure of the average distance between atoms of superimposed proteins. Due to extremely low Rmsd values The largest difference can be seen in monomer B, which is to be expected considering the oxygen bound to monomer B causing some structural differences.

Table 5.13 Rmsd values for the differences between peracetic acid and ferric structures of DHP-B calculated using 'align' on PyMOL (Hart et al 2015)

Measured	Value (Å)
Rmsd Overall	0.006
Rmsd monomer A	0.001
Rmsd monomer B	0.024

At haem B, the porphyrin is seen as 'domed', and the iron is pulled out of plane, with iron being at a 95° angle. Normally, the porphyrin is closer to 180°. This is due to the iron interacting with the NH group of the distal histidine (His55), but this could also be due to the bound oxygen interacting with the NH. This is common in haem proteins without ligands bound, but also in the intermediate 'compound I', as the His55 is pulled closer to the haem plane to interact with the oxygen.

5.6.3 Oxyferrous DHP-B

The data for oxyferrous DHP-B was measured in SACL, Japan by Dr Tadeo Moreno-Chicano in 2018. The data was then processed by me in the same way outlined for the DHP-B peracetic acid soak outlined in 5.3.2.

The overall SFX structure at 2 Å of oxyferrous DHP appeared as expected, again crystallising as a dimer in the asymmetric unit. The globin fold is also apparent, with seven α -helices per monomer.

The F_o-F_c map for oxyferrous DHP (Figure 5.17) shows a clear positive electron density above the haem plane in both the monomers. Oxygen molecules were modelled in to a small cloud of positive difference density and refined using REFMAC5 but only at haem B did the electron density support the oxygen, so was removed from haem A. However, haem A shows a hemichrome species, with a double conformation of the haem cofactor. Previously, the existence of a hemichrome species was to do with the use of glycerol as a cryoprotectant for conventional x-ray crystallography. With these species being present in SFX structures, where no cryoprotectant is used, this can be ruled out. However, it could be a crystallisation artefact from being in contact with PEG for long periods of time (Mao *et al.*, 2013) or have biological relevance. To be able to confirm whether it has a biological relevance, further investigation needs to be done.

5.6.4. Comparison between SFX room temperature and cryostructure oxyferrous DHP-B

Oxyferrous DHP-B outlined in this report was measured at room temperature, at an XFEL site to produce a damage-free structure. Here, there will be a comparison made between the cryogenic and room temperature structures. The cryogenic DHP-B structure was obtained from the pdb (5V5J) at 1.8 Å. An overall superposition of both of these structures is shown in Figure 5.23. Comparing the two structures is important to see whether the temperature at which the crystals are measured at has an effect on the structure. Cryogenic temperatures, although give a higher resolution due to reducing radiation damage, could mask the structure *in vivo* and therefore leave out potentially biologically relevant conformations.

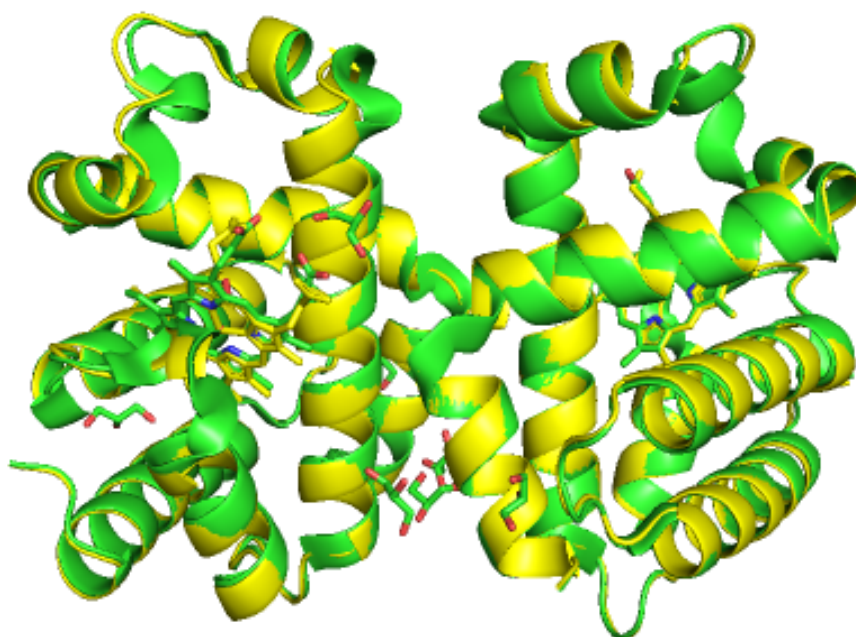


Figure 5.22. Superposition of DHP-B Oxyferrous measured at room temperature (yellow) and cryogenic temperatures (green) Monomer A is shown on the left and monomer B on the right.

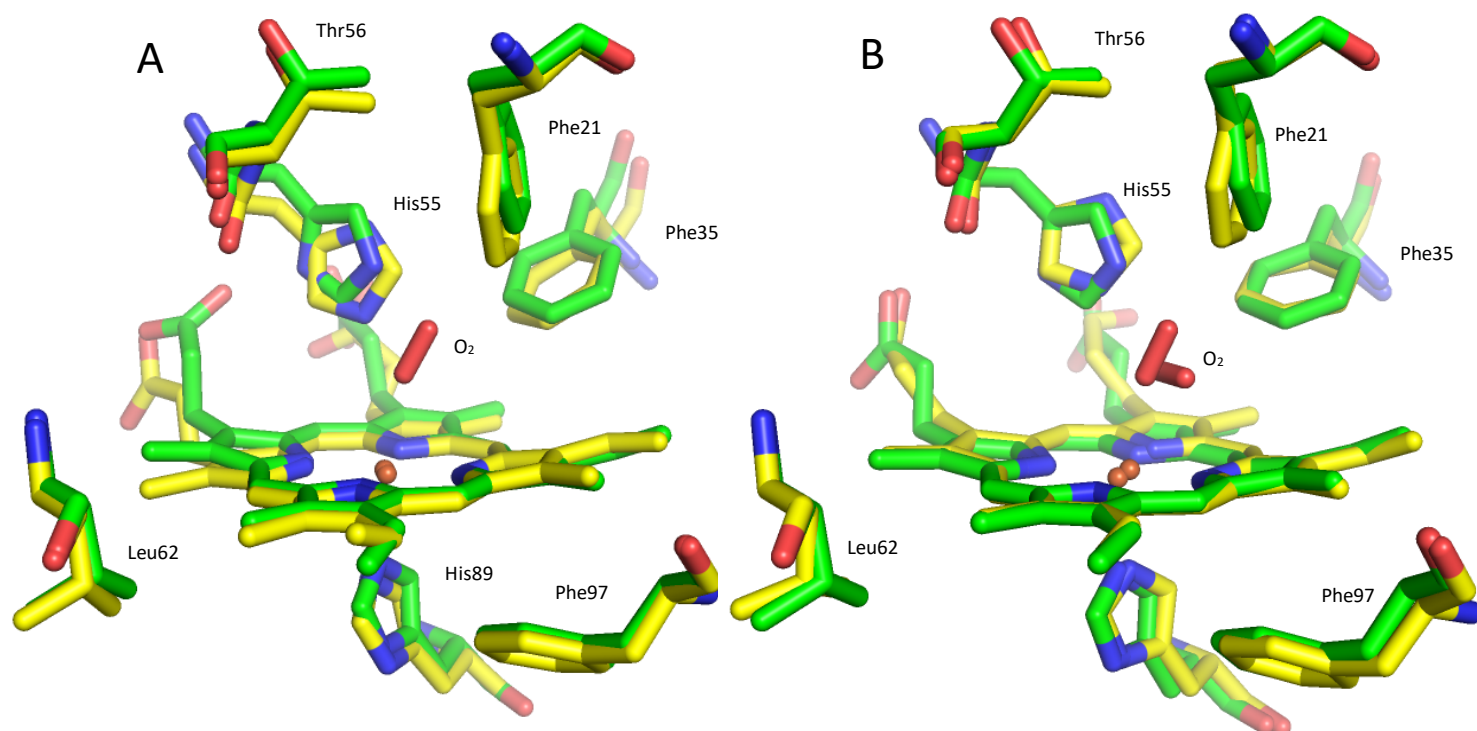


Figure 5.23. Superposition of DHP-B Oxyferrous measured at room temperature (yellow) and cryogenic temperatures (green). Haems A and B are labelled respectively.

The structures are very similar, with only some differences in the overall structure being visible, close to the haem cofactors mainly in monomer B. The Rmsd values between the two structures are shown in Table 5.7.

Table 5.14 Rmsd values for the differences between peracetic acid and ferric structures of DHP-B calculated using 'align' on PyMOL (Hart et al 2015)

Measured	Value (Å)
Rmsd Overall	0.313
Rmsd monomer A	0.257
Rmsd monomer B	0.306

The Rmsd values agree with the observation that the two structures are similar, as they are relatively small Rmsd values. The differences around the haem groups could be due to the way in which the oxygen is bound, as the oxygen in the RT oxyferrous structure is almost parallel to the haem plane. The cryogenic structure of DHP-B haem B has the oxygen bound at an approximate 113° angle, this is close to the energetically favourable 115° bond seen in haem proteins, as the oxygen is sp² hybridized when bound to haem iron. This suggests that the temperature in which the structure is measured has an effect on the way that oxygen binds to DHP-B, the reasoning for this would need further investigation. In both the SFX and cryogenic structures the oxygen forms hydrogen bonds with the N² atom of the distal His55, with distances of 3.41 Å and 3.24 Å respectively.

Chapter 6 - General Discussion

6.1 Introduction

Dehaloperoxidase is a multifunctional globin, functioning as a peroxidase, haemoglobin, peroxygenase, oxidase and oxygenase. DHP has two isoforms, DHP-A and DHP-B, with DHP-B being the focus of this thesis. DHP-B has 96% amino acid identity with its isoform, but has mechanistic differences including having a fourfold higher peroxidase reactivity. DHP is highly unusual in that it has so many biologically relevant functions, unlike most haem-proteins that have one key function. What also sets DHP apart from other globins like myoglobin, is the distal histidine (His55), located at the opening of the distal cavity. This histidine is key to DHP's peroxidase function, with an 'open' and 'closed' conformation. The distal histidine is also the only polar amino acid present in the binding pocket, unlike other peroxidases.

His55 is key to ligand binding, being 'swung out' of the pocket into its open conformation to allow room for organic substrates to bind. When His55 is 'closed', it can interact with the sixth ligand of the haem iron. In Chapter 3, binding of different aromatic ligands are compared from literature and show the 'open' conformation of His55. All of these ligands bind differently to DHP-B, showing how varied ligand binding can be in DHP-B, relating back to DHP's ability to have many different functions, as well as its ability to switch between these different functions, which is currently unknown.

One of the ways in which ligand binding to DHP has been investigated in this thesis is using serial crystallography. Serial crystallography uses small microcrystals and is measured at room temperature. This gives damage-free structures, and also potentially removes any bias created from using cryogenic temperatures or using cryoprotectants (Fraser *et al.*, 2011). Serial crystallography was performed at both XFELs (SACLA, Japan) and synchrotrons (Diamond Light Source, UK).

6.2 Serial Crystallography

The comparison of room temperature and cryogenic of oxyferrous DHP-B revealed some key differences. Cryogenic temperatures, although give a higher resolution, can mask the *in vivo* structure, as well as leaving out biologically relevant conformations. With the oxyferrous structures being overall quite similar, the main focus was on the binding pocket itself. With haem B (from monomer B) for both oxyferrous structures having oxygen bound, the focus was mainly on this haem. The oxygen bound in the cryogenic temperature structure was at a more energetically favoured angle then in the room temperature structure. There were also differences in the angle of Phe35, one of the hydrophobic residues in the distal pocket that may be important for ligand binding. The oxygen in the room temperature structure were also closer, with His55 being even closer to the ligand, supporting the previously stated function of His55 to bind to the six-coordinated ligand bound to the Fe of the haem. The oxygen is also within binding distance of the iron, being 2Å from the iron.

These are some small differences in the structures but show importantly how room temperature studies can give different results. More room temperature studies would need to be compared to give a better insight to how much temperature can affect the structure.

6.3 Limitations of this study

One major limitation of this study was the production of microcrystals for investigation. Large amounts of microcrystals were needed for the serial data collections, which was a problem very difficult to overcome. With each protein preparation, optimization was needed with many different variables. The production of microcrystals was unsuccessful during this project, with this issue being overcome with large protein preparations, as well as more time to do the time-consuming optimisation task after each prep.

Another limitation was the ligand soaking conditions. Ligand soaks were

unsuccessful due to too low concentrations used, due to the limited availability of crystals and the possibility of damage to these due to the nature of these soaks. More trials would need to be set up, with both single crystal studies and microcrystals to find an optimal soaking concentration to be able to see these ligands bound in the distal pocket.

The DHP oxyferrous data showed that the microcrystals used were twinned. This presented challenges with processing and refining the data. CCP4i2 was used with this data, as *REFMAC5* worked with the twinned data better. After, the low-resolution pipeline *LORESTR* was used. This is recommended for twinned data, and automatically determines whether the data is twinned or not and runs 40 cycles of refinement, using homologous structures from the pdb to provide external restrains (Kovalevskiy, Nicholls and Murshudov, 2016)

Bibliography

- Afonine, P. V. *et al.* (2012) 'Towards automated crystallographic structure refinement with phenix.refine', *Acta Crystallographica Section D: Biological Crystallography*, 68(4), pp. 352–367. doi: 10.1107/S0907444912001308.
- Bailly, X. *et al.* (2007) 'Globin gene family evolution and functional diversification in annelids', *FEBS Journal*, 274(10), pp. 2641–2652. doi: 10.1111/j.1742-4658.2007.05799.x.
- Barrios, D. A. *et al.* (2014) 'Peroxygenase and Oxidase Activities of Dehaloperoxidase-Hemoglobin from *Amphitrite ornata*', *Journal of the American Chemical Society*, 136(22), pp. 7914–7925. doi: 10.1021/ja500293c.
- Barty, A. *et al.* (2019) 'Summary for Policymakers', in Intergovernmental Panel on Climate Change (ed.) *Climate Change 2013 - The Physical Science Basis*. Cambridge: Cambridge University Press, pp. 1–30. doi: 10.1017/CBO9781107415324.004.
- Berman, H. M. *et al.* (2002) 'The Protein Data Bank', *Acta Crystallographica Section D Biological Crystallography*, 58(6), pp. 899–907. doi: 10.1107/S0907444902003451.
- Boutet, S. *et al.* (2012) 'High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography', *Science*, 337(6092), pp. 362–364. doi: 10.1126/science.1217737.
- Carey, L. M. *et al.* (2018) 'How nature tunes isoenzyme activity in the multifunctional catalytic globin dehaloperoxidase from *Amphitrite ornata*', *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1866(2), pp. 230–241. doi: 10.1016/j.bbapap.2017.11.004.
- Chapman, H. N., Coleman, C. and Timneanu, N. (2014) 'Diffraction before destruction', *Philosophical Transactions of the Royal Society B: Biological Sciences*,

369(1647), p. 20130313. doi: 10.1098/rstb.2013.0313.

Chen, Y. P. *et al.* (1991) 'Purification and properties of a unique flavin-containing chloroperoxidase from the capitellid polychaete *Notomastus iobatus*', *Journal of Biological Chemistry*, 266, pp. 23909–23915.

Chen, Z. *et al.* (2009) 'Distal histidine conformational flexibility in dehaloperoxidase from *Amphitrite ornata*', *Acta Crystallographica Section D Biological Crystallography*, 65(1), pp. 34–40. doi: 10.1107/S0907444908036548.

Chinte, B. *et al.* (2007) 'Cryogenic (<20K) helium cooling mitigates radiation damage to protein crystals', *Acta Crystallographica Section D: Biological Crystallography*, 63, pp. 486–492.

Cowtan, K. (2006) 'The Buccaneer software for automated model building. 1. Tracing protein chains', *Acta Crystallographica Section D Biological Crystallography*, 62(9), pp. 1002–1011. doi: 10.1107/S0907444906022116.

Cusack, S. *et al.* (1998) 'Small is beautiful: Protein micro-crystallography', *Nature Structural & Molecular Biology*, 5, pp. 634–637.

D'Antonio, J., D'Antonio, E., Thompson, M.K., Bowden, E.F., Franzen, S., Smirnova, T. and Ghiladi, R. A. (2010) 'Spectroscopic and Mechanistic Investigations of Dehaloperoxidase B from *Amphitrite ornata*', *Biochemistry*, 66, pp. 529–538.

D'Antonio, J. *et al.* (2010) 'Spectroscopic and Mechanistic Investigations of Dehaloperoxidase B from *Amphitrite ornata*', *Biochemistry*, 66, pp. 529–538.

Du, J. *et al.* (2011) 'Amphitrite ornata Dehaloperoxidase (DHP): Investigations of Structural Factors that Influence the mechanism of halophenol dehalogenation using "peroxidase-like" myoglobin mutants and "myoglobin-like" DHP mutants', *Biochemistry*, 50, pp. 8172–8180.

Emsley, P. *et al.* (2010) 'Features and development of Coot', *Acta Crystallographica Section D: Biological Crystallography*, 66, pp. 486–501.

Feducia, J. *et al.* (2009) 'Characterization of dehaloperoxidase compound ES and its reactivity with trihalophenols', *Biochemistry*, 48, pp. 995–1005.

Francesco, N. *et al.* (2010) 'New Insights into the role of distal histidine flexibility in ligand stabilization of Dehaloperoxidase-Hemoglobin from *Amphitrite ornata*', *Biochemistry*, 49, pp. 1903–1912.

Franzen, S. *et al.* (2016) *Multi-functional haemoglobin dehaloperoxidases' in Raven, E. and Dunford, B.* Royal Society of Chemistry.

Fraser, J. S. *et al.* (2009) 'Hidden alternative structure of proline isomerase essential for catalysis', *Nature*, 462, pp. 669–673.

Fraser, J. S. *et al.* (2011) 'Accessing protein conformational ensembles using room-temperature X-ray crystallography', *Proceedings of the National Academy of Sciences of the United States of America*, 108, pp. 16247–16252.

Garman, E. F. (2010) 'Radiation damage in macromolecular crystallography: what is it and why should we care?', *Acta Crystallographica Section D Biological Crystallography*, 66(4), pp. 339–351. doi: 10.1107/S0907444910008656.

Gati, C. *et al.* (2013) 'Serial Crystallography on in vivo grown microcrystals using synchrotron radiation', *IUCrJ*, 1, pp. 87–94.

Girvan, G. M. and Munro, A. W. (2013) 'Heme Sensor Proteins', *Journal of Biological Chemistry*, 288, pp. 13194–13203.

Grimes, J. M. *et al.* (2018) 'Where is crystallography going?', *Acta Crystallographica Section D Structural Biology*, 74(2), pp. 152–166. doi: 10.1107/S2059798317016709.

Guha, S. *et al.* (2012) 'Fabrication of X-ray compatible microfluidic platforms for protein crystallization', *Sensors and Actuators B: Chemical*, 174, pp. 1–9.

Hampel, A. *et al.* (1968) 'Single crystals of transfer RNA from formylmethionine and phenylalanine transfer RNAs', *Science*, 162, pp. 1384–1387.

Hardison, R. (1998) 'Hemoglobins from bacteria to man: evolution of different patterns of gene expression', *Journal of Experimental Biology*, pp. 1099–1117.

Kovalevskiy, O., Nicholls, R. A. and Murshudov, G. N. (2016) 'Automated refinement of macromolecular structures at low resolution using prior information', *Acta Crystallographica Section D Structural Biology*, 72(10), pp. 1149–1161. doi: 10.1107/S2059798316014534.

LaCount, M. W. *et al.* (2000) 'The Crystal structure and amino acid sequence of dehaloperoxidase from *Amphitrite ornata* indicate common ancestry with globins', *Journal of Biological Chemistry*, 275, pp. 18712–18716.

Lebioda, L. *et al.* (1999) 'An Enzymatic Globin from a Marine Worm', *Nature*, 401, p. 445.

Lincoln, D. E. *et al.* (2005) 'Bromophenol Accumulation and Sediment Contamination by Marine Annelids *Notomastus lobatus* and *Thelepus crispus*', *Biochemical Systematics and Ecology*, 6, pp. 559–570.

Luft, J. R. *et al.* (1994) 'A macromolecular crystallization procedure employing diffusion cells of varying depths as reservoirs to tailor the time course of equilibration in hanging drop and sitting drop vapour diffusion and microdialysis experiments', *Journal of Applied Crystallography*, 27, pp. 443–452.

Malewschik, T. *et al.* (2019) 'The multifunctional globin dehaloperoxidase strikes again: Simultaneous peroxidase and peroxygenase mechanisms in the oxidation of EPA pollutants', *Archives of Biochemistry and Biophysics*, 673, p. 108079. doi:

10.1016/j.abb.2019.108079.

Mao, L. *et al.* (2013) 'Horseradish Peroxidase Inactivation: Heme Destruction and Influence of Polyethylene Glycol', *Scientific Reports*, 3(1), p. 3126. doi: 10.1038/srep03126.

McCombs, N. L. *et al.* (2016) 'Nonmicrobial nitrophenol degradation via peroxygenase activity of dehaloperoxidase-haemoglobi from *Amphitrite ornata*', *Biochemistry*, 55, pp. 2465–2478.

McCoy, A. J. *et al.* (2007) 'Phaser crystallographic software', *Journal of Applied Crystallography*, 40(4), pp. 658–674. doi: 10.1107/S0021889807021206.

McPherson, A. (1999) *Crystallization of Biological Macromolecules*. New York: John Wiley and Sons.

Meents, A. *et al.* (2010) 'Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures', *Proceedings of the National Academy of Sciences of the United States of America*, 107, pp. 1094–1099.

Moody, P. C. E. and Raven, E. L. (2018) 'The Nature and Reactivity of Ferryl Heme in Compounds I and II', *Accounts of Chemical Research*, 2, pp. 427–435.

Moreno-Chicano, T. *et al.* (2019) 'High-throughput structures of protein–ligand complexes at room temperature using serial femtosecond crystallography', *IUCrJ*, 6(6), pp. 1074–1085. doi: 10.1107/S2052252519011655.

Morita, Y. *et al.* (1988) 'Purification, Crystallization, and Characterization of Peroxidase from *Coprinus cinereus*', *The Journal of Biochemistry*, 103(4), pp. 693–699. doi: 10.1093/oxfordjournals.jbchem.a122331.

Mueller, C. *et al.* (2015) 'Fixed target matrix for femtosecond time-resolved and in

situ serial micro-crystallography', *Structural Dynamics*, 2(5), p. 054302. doi: 10.1063/1.4928706.

Murray, J. and Garman, E. F. (2002) 'Investigation of possible free-radical scavengers and metrics for radiation damage in protein cryocrystallography', *Journal of Synchrotron Radiation*, 9(6), pp. 347–354. doi: 10.1107/S0909049502014632.

Murshudov, G. N. *et al.* (2011) 'REFMAC 5 for the refinement of macromolecular crystal structures', *Acta Crystallographica Section D Biological Crystallography*, 67(4), pp. 355–367. doi: 10.1107/S0907444911001314.

Neutze, R. *et al.* (2000) 'Potential for biomolecular imaging with femtosecond X-ray pulses', *Nature*, 406, pp. 753–757.

Neyadi, A. A., Rauf, M. A. and Ashraf, S. S. (2018) 'Oxidoreductases for the remediation of organic pollutants in water – a critical review', *Critical Reviews in Biotechnology*, 38, pp. 1–18.

Nicoletti, F. P. *et al.* (2011) 'Degradation of sulfide by dehaloperoxidase-hemoglobin from *Amphitrite ornata*', *Journal of Biological Inorganic Chemistry*, 16(4), pp. 611–619.

Ortiz de Montellano, P. R. (1987) 'Control of the catalytic activity of prosthetic heme by the structure of hemoproteins', *Accounts of Chemical Research*, 20(8), pp. 289–294. doi: 10.1021/ar00140a004.

Osborne, R. L. *et al.* (2006) 'C. fumago chloroperoxidase is also a dehaloperoxidase: oxidative dehalogenation of Halophenols', *Journal of American Chemical Society*, 128, pp. 1036–1037.

Owen, R. L. *et al.* (2017) 'Low-dose fixed-target serial synchrotron crystallography', *Acta Crystallographica Section D: Structural biology*, 73, pp. 373–378.

Owen, R. L., Rudino-Pinera, E. and Garman, E. F. (2006) 'Experimental determination of the radiation dose limit for cryocooled protein crystals', *Proceedings of the National Academy of Sciences*, 103(13), pp. 4912–4917. doi: 10.1073/pnas.0600973103.

Paoli, M., Marles-Wright, J. and Smith, A. (2002) 'Structure-function relationships in heme-proteins', *DNA Cell Biology*, 4, pp. 271–280.

Plummer, A., Matthew, T. and Franzen, S. (2013) 'Role of Polarity of the Distal Pocket in the Control of inhibitor binding in Dehaloperoxidase-haemoglobin', *Biochemistry*, 52, pp. 2218–2227.

Poulos, T. L. (2014) 'Heme Enzyme Structure and Function', *Chemical Reviews*, 114(7), pp. 3919–3962. doi: 10.1021/cr400415k.

Reedy, C. J., Elvekrog, M. M. and Gibney, B. . (2008) 'Development of a heme protein structure-electrochemical function database', *Nucleic Acids Research*, 36, pp. 307–313.

Schneider, S. *et al.* (2007) 'Diversity and Conservation of Interactions for binding of haem proteins', *Natural Product Reports*, 24, pp. 621–630.

de Serrano, V. *et al.* (2007) 'X-ray crystal structural analysis of the binding site in the ferric and oxyferric forms of the recombinant heme dehaloperoxidase cloned from *Amphitrite ornata*', *Acta Crystallographica Section D: Biological Crystallography*, 63, pp. 1094–1101.

de Serrano, V. *et al.* (2010) 'Structure of dehaloperoxidase B at 1.58 Å resolution and structural characterization of the AB dimer from *Amphitrite ornata*', *Acta Crystallographica Section D Biological Crystallography*, 66(5), pp. 529–538. doi: 10.1107/S0907444910004580.

- Sivaraja, M. *et al.* (1989) 'Identification by ENDOR of Trp191 as the free radical site in cytochrome c peroxidase compound ES', *Science*, 245, pp. 738–740.
- Smith, J. L., Fischetti, R. F. and Yamamoto, M. (2012) 'micro-crystallography comes of age', *Current opinions in Structural Biology*, 22, pp. 602–612.
- Thompson, M. K. *et al.* (2011) 'Dehaloperoxidase-Hemoglobin from *Amphitrite ornata* Is Primarily a Monomer in Solution', *The Journal of Physical Chemistry B*, 115(14), pp. 4266–4272. doi: 10.1021/jp201156r.
- Uervirojnangkoorn, M. *et al.* (2015) 'Enabling X-ray free electron laser crystallography for challenging biological systems from a limited number of crystals', *eLife*, 4. doi: 10.7554/eLife.05421.
- Vergara, A. *et al.* (2009) 'Correlation between Hemichrome Stability and the Root Effect in Tetrameric Hemoglobins', *Biophysical Journal*, 97(3), pp. 866–874. doi: 10.1016/j.bpj.2009.04.056.
- Vlasits, J. *et al.* (2010) 'Mechanisms of catalase activity of heme peroxidases', *Archives of Biochemistry and Biophysics*, 500(1), pp. 74–81. doi: 10.1016/j.abb.2010.04.018.
- Warkentin, M. *et al.* (2013) 'Global radiation damage: temperature dependence, time dependence and how to outrun it', *Journal of Synchrotron Radiation*, 20(1), pp. 7–13. doi: 10.1107/S0909049512048303.
- Weber, R. E. *et al.* (1977) 'Hemoglobins of two terebellid polychaetes: *Enoplobranchus sanguineus* and *Amphitrite ornata*', *Comparative Biochemistry and Physiology Part A: Physiology*, 56(2), pp. 179–187. doi: 10.1016/0300-9629(77)90182-7.
- White, T. A. *et al.* (2012) 'CrystFEL: a software suite for snapshot serial

crystallography', *Journal of Applied Crystallography*, 45, pp. 335–341.

Yuan, Z., Bailey, T. L. and Teasdale, R. D. (2005) 'Prediction of protein B-factor profiles', *Proteins: Structure, Function, and Bioinformatics*, 58(4), pp. 905–912. doi: 10.1002/prot.20375.

Zhang, E. *et al.* (1996) 'Crystallization and initial spectroscopic characterization of the heme-containing dehaloperoxidase from the marine polychaete *Amphitrite ornata*', *Acta Crystallographica Section D Biological Crystallography*, 52(6), pp. 1191–1193. doi: 10.1107/S09074444996007974.

Zwart, P. H., Grozse-Kunstleve, R. W. and Adams, P. D. (2005) *Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation*. Available at: <https://bit.ly/2VWX888>.